EXHIBIT A

UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

----------x

AUTHORS GUILD, *et al.*,

Plaintiffs,

v. 23 Civ. 8292 (SHS)(OTW)

OPENAI, INC., *et al.*,

Conference

Defendants.

----------x

JULIAN SANCTON, *et al.*,

Plaintiffs,

v. 23 Civ. 10211 (SHS)(OTW)

OPENAI, INC., *et al.*,

Defendants.

----------x

THE NEW YORK TIMES COMPANY,

Plaintiff,

v. 23 Civ. 11195 (SHS)(OTW)

MICROSOFT CORPORATION, et al.,

Defendants.

----------x

----------x

NICHOLAS A. BASBANES, et al.,

Plaintiffs,

v. 24 Civ. 00084 (SHS)(OTW)

MICROSOFT CORPORATION, et al.,

Defendants.

----------x

APPEARANCES

SUSMAN GODFREY LLP
Interim Class Counsel for Authors Guild and Alter Class Actions
BY: ROHIT NATH

SUSMAN GODFREY LLP
Attorneys for The New York Times
BY: KATHERINE PEASLEE
ZACH SAVAGE
ALEXANDER P. FRAWLEY
ADNAN MUTTALIB
IAN CROSBY
DEMETRI BLAISDELL

ROTHWELL FIGG
Attorneys for New York Times and Daily News
BY: JENNIFER MAISEL
STEVEN LIEBERMAN
JENNY COLGATE

LOEVY & LOEVY
Attorneys for Center for Investigative Reporting
BY: MATTHEW TOPIC

LIEFF CABRASER HEIMANN & BERNSTEIN, LLP
Attorneys for Authors Guild and Alter Class Plaintiffs
BY: RACHEL GEMAN

APPEARANCES
(Continued)

KEKER VAN NEST & PETERS
Attorneys for OpenAI
BY: JAMES SLAUGHTER
MICHELLE YBARRA
ANDREW DAWSON
CHRISTOPHER SUN
SARAH SALOMON

MORRISON & FOERSTER LLP
Attorneys for OpenAI
BY: JOSEPH GRATZ
CAROLYN HOMER

LATHAM & WATKINS LLP
Attorneys for Microsoft
BY: ELANA NIGHTINGALE DAWSON

FAEGRE DRINKER BIDDLE & REATH LLP
Attorneys for Microsoft
BY: JARED BRIANT

ORRICK HERRINGTON & SUTCLIFFE LLP
Attorneys for Microsoft Corporation
BY: ANNETTE L. HURST
SHERYL GARKO

ALSO PRESENT:
KAREN CHESLEY, The New York Times
NICK STANDISH, The New York Times

(Cases called)

THE DEPUTY CLERK: Please state your appearances for the record.

MR. CROSBY: Ian Crosby from Susman Godfrey for The New York Times.

MR. NATH: Good morning, your Honor. Rohit Nath from Susman Godfrey for the class plaintiffs.

MR. SAVAGE: Good morning, your Honor. Zach Savage from Susman Godfrey for The New York Times.

MS. MAISEL: Good morning, your Honor. Jennifer Maisel from Rothwell Figg on behalf of The New York Times and Daily News.

MS. GEMAN: Good morning, your Honor. Rachel Geman, Lieff Cabraser, class plaintiffs.

MR. TOPIC: Matt Topic, Loevy & Loevy, Center for Investigative Reporting.

THE COURT: All right.

JUDGE BAHRAOUI: Good morning, your Honor. Adnan Muttalib for the New York Times.

MS. PEASLEE: Good morning, your Honor. Katy Peaslee from Susman Godfrey on behalf of The New York Times.

MS. COLGATE: Good morning, your Honor. Jenny Colgate from Rothwell Figg on behalf of the Daily News.

MR. LIEBERMAN: Good morning, your Honor. Steve Lieberman, Rothwell Figg, on behalf of The New York Times

Company and the Daily News plaintiffs.

MR. FRAWLEY: Alexander Frawley from Susman Godfrey on behalf of The New York Times.

MS. CHESLEY: None of us are speaking.

THE COURT: OK. Welcome anyway.

MR. SLAUGHTER: Good morning, James Slaughter, Keker Van Nest & Peters, on behalf of OpenAI.

MS. YBARRA: Good morning, your Honor. Michelle Ybarra, Keker Van Nest & Peters, also on behalf of OpenAI.

MR. HURST: Good morning, your Honor. Annette Hurst from Orrick on behalf of Microsoft.

MS. GARKO: Good morning, your Honor. Sheryl Garko from Orrick on behalf of Microsoft.

MR. DAWSON: Andrew Dawson from Keker Van Nest & Peters on behalf of OpenAI.

MS. NIGHTINGALE DAWSON: Good morning, your Honor. Elana Nightingale Dawson from Latham Watkins on behalf of OpenAI.

MR. BRIANT: Good morning, your Honor. Jared Briant, Faegre Drinker, on behalf of Microsoft.

MR. BAILEY: Good morning, your Honor. Edward Bailey of Keker Van Nest & Peters on behalf of OpenAI.

MS. SOLOMON: Good morning, your Honor. Sarah Solomon, also from Keker Van Nest & Peters, on behalf of OpenAI.

MR. SUN: Good morning, your Honor. Christopher Sun, Keker Van Nest & Peters, also on behalf of OpenAI.

MS. HOMER: Good morning, your Honor. Carolyn Homer from Morrison & Foerster on behalf of OpenAI.

THE COURT: All right. Good morning.

All right. I might end up messing up your seating because we're going to take the issues in the news cases first and we're going to actually start -- we're not going to go in order on the chart, although I thank you for the chart with the hyper links. we are going to start with the AEO designation issue.

This is styled, I think, as a motion to compel from the plaintiffs, right?

MR. CROSBY: I think substantively what we're asking for is a modification of the protective order. We have been trying very hard to avoid having to seek to compel Microsoft to re-designate its production, notwithstanding the fact that we are still at 60 percent, very close to the presumptively abusive 70 percent AEO designation.

We managed to work around that designation for purposes of sharing certain documents with our clients. We still just have a problem with using documents in depositions with OpenAI. The issue is Microsoft and OpenAI were collaborating on the products at issue here, and your Honor has set aside discovery on Microsoft's new products that are

working on separately from OpenAI. So that issue is not really there.

And the problem is, of course, ineffective cross-examination is greatly assisted by the element of surprise and by having receipts. And because of the way that Microsoft has designated its production, we are really at a disadvantage with respect to both of those elements with using documents in the depositions.

Further, there is a separate issue about the deposition protocol, where at least the other side is trying to greatly restrict the hours we have.

THE COURT: We are not talking about that until last.

MR. CROSBY: So, but having to this Mother May I with AEO-designated documents is going to only make that time more precious.

That's the issue, your Honor.

THE COURT: OK. I hear you.

Just a minute. I did review the samples that I think were attached to ECF 386 or 368, I can't remember the exact number.

And on the one hand, I question a little bit -- I'm not going to get into the details of what is contained in those documents because I know they are under seal -- but they didn't initially strike me as documents that ought to have an AEO designation.

At the same time, I get what you're saying about the Mother May I game, whether that needs to happen at a deposition or that can happen before. I believe it was Microsoft -- and maybe this is what Ms. Hurst was going to say when she popped up -- that there was -- that they did propose a protocol that happens pretty commonly which is to discuss some of the exhibits in advance.

MR. CROSBY: Well, I mean, there goes the element of surprise, right?

You know --

THE COURT: Well, yeah. I mean, although I'm not going to tell you how to litigate a trial or how to take a deposition.

But yes, there goes some of the element of surprise. At the same time, I mean, I remember back when I was doing this and we had to have binders made rather than electronic documents, huge binders. I mean ...

MR. CROSBY: Your Honor, I've never given a witness exhibits in advance of a cross-examination in my life as a trial lawyer. It's just not how it's done, if you want to do it effectively, in my personal view.

Certainly, it might be a protocol that could be reasonable if they designated 10 percent of their documents as AEO or 20 percent of their documents as AEO. It becomes completely intractable, basically every hot document in the

case.

As your Honor pointed out, I don't understand the rationale why the four examples that we gave you just now, exposed four more pieces of work product. And do we have to do that for every single document that we want to use in a deposition.

Again, if they would re-designate their production and if this court were to say your hard cap i you have 10 percent or 15 percent or 20 percent, use it wisely, that would be another circumstance. We don't want to make them do that. It would be -- I'm sure it would cost a lot of money, take a lot of lawyer time that could be better used on the merits of the case.

We are just trying to get around this over-designation so that we can effectively use these documents in depositions without too many administrative difficulties. The solution we proposed is just simply that the witness will -- first of all, it has to be reasonably necessary, so we're not going to be just taking random documents out of the blue that have nothing to do with a particular witness. What is the point in that? Is it reasonably necessary.

And then they have to do the protective order undertaking, which is I will only use these, among other things, that says I'll only use these for purposes of the lawsuit. So, again, these documents, they concern the receipts

with respect to the collaboration between OpenAI and Microsoft. Typically or often they are Microsoft's documents saying I had this interaction with somebody at OpenAI.

THE COURT: Well, that actually brings me to the next issue, which is whether we're creating a problem for ourselves in the sense that, the way I read the language in the protective order, it's limited to author recipient custodian or other person who otherwise possessed or knew the information.

I mean, doesn't that -- I mean, it seemed to me that the proposed examples that you provided in, I believe it's ECF 386 --

MR. CROSBY: Yes.

THE COURT: -- seems like those documents would be covered under that "or other person who has knowledge of the information."

MR. CROSBY: I think if we could get, sort of, a safe harbor that if on the face of the document, it looks like somebody knew or should have known that information or is referring to an interaction, I think that would get us a long way.

I think there was one document where somebody from Microsoft -- and I'll be very vague because we're in open court -- basically saying, if we don't deliver on this, it would be a problem for OpenAI. And I would like to ask the OpenAI witness, did they deliver on that, was that a problem

for you. If a document refers back and forth to their collaboration, if that is fair game, I think that maybe we could live with that interpretation.

But my concern is that a lawyer on the other side during a deposition will say, well, they didn't know about the exact contents of this document or everything in this document, even though it refers to subject matter of which they would have common knowledge. If we can make sure we have a safe harbor and we're not going to get in trouble when a document on its face is referring to something a witness should know --

THE COURT: I get the sense that, you know, a lot of this is trying to frontload, because you're talking about deposition protocols and potential limits on hours, that things can go awry that way, and time can be spent making objections about documents and having discussions about documents that seem to be not appropriate.

If that happens, you will get more time. OK. If I have to be called upon to make rulings about how, whether these were objections made or make rulings about them at the point of the deposition, believe me, there will be consequences. Because the way that we have been talking about these documents, the examples that are set, seems to me -- I mean, it's not an advisory ruling, but I guess it's just that -- it seems to me that there is room in the protective order already.

And I'm hesitant to go wholesale and say, look, I also

think that 60 percent of all the documents produced being marked AEO is probably a little bit of an over-designation. That said, I'm not sure we're at the point where I want to say, go back and go re-designate all of these documents. I wonder if there is something that can be done in between that can work.

Ms. Hurst, I cut you off earlier. Let me hear from you.

MR. HURST: Thank you, your Honor. Annette Hurst from Microsoft.

I think the court is exactly right, that if there is clearly on the face of the document knowledge of the information, the examining attorney can ask about that. However, what they are proposing is to completely eliminate the AEO designation, because that is what the effect of this would be.

Let's just think about how this would work. There is no element of surprise because the witness can't be shown the document unless the witness agrees to sign the undertaking to the protective order. But how would an individual witness from OpenAI ever know whether they could appropriately undertake that obligation, frankly, without the advice of personal counsel at the deposition and without looking at the document first. The idea that somebody would agree to sign off on a new confidentiality obligation to another company altogether in the

middle of a deposition makes no sense.

THE COURT: Wait a minute. Wait a minute. Back up a little bit.

What are you talking about here about signing an undertaking?

That normally happens when you're taking a deposition and being shown documents with a protective order.

MR. HURST: Not for the witnesses, your Honor.

The witnesses are ordinarily shown documents that they have a foundation for, so the very first principle here is that they are seeking to show properly designated AEO documents to witnesses who have no foundation for that, who have never seen them before, who may not have any knowledge whatsoever of the contents.

Now, they have cherrypicked a few examples where the witnesses might have some knowledge of the issues under discussion, but the vast majority of the documents aren't like that and they don't have any dispute about whether they are properly designated AEO. And they nevertheless want the option to show documents to witnesses with no foundation for them whatsoever, purely internal Microsoft documents, to a competitor, OpenAI.

THE COURT: Do you have a joint defense agreement in place?

MR. HURST: We do have a joint defense agreement?

THE COURT: What does it say about document sharing or internal documents?

MR. HURST: I don't remember exactly what it says, but I can tell you, by and large, we do not do that, your Honor.

These companies are competitors. OpenAI announced a new joint venture with Oracle and SoftBank this morning, your Honor. The companies behave appropriately as competitors and do not share sensitive technical and business information with one another in light of that status.

And it would not be appropriate for plaintiffs' counsel to come and show internal Microsoft documents to its competitor, OpenAI. The individual witnesses have no confidentiality obligation to Microsoft, nor could they agree willingly to undertake such an obligation in the course of a deposition simply to facilitate plaintiffs' questioning on documents where the witnesses completely lack foundation.

I mean, if we just think about it, a document the witness has never seen before, right. They have never seen it. Imagine it discusses a very confidential or sensitive Microsoft piece of business strategy related to this market. And Mr. Crosby says, Well, I want to play gotcha on cross by showing them Microsoft's plan here to a witness who has never seen this before.

And then he's go going to stop and he's going to ask counsel for OpenAI, Will the witness agree to sign the

undertaking to the protective order so I can show the witness a confidential Microsoft document?

It's just a recipe for delay, inefficiency, and waste because the witness has the right to say no to that. And how is the witness going to make that examination and make that decision for him or herself without seeing the document first?

Now --

THE COURT: Let me just stop.

We started this out where I set aside new products, right. We are talking about old products and old joint work between OpenAI and Microsoft, right?

MR. HURST: Well, your Honor, certainly we are not doing discovery about the new products under development, but that doesn't mean we haven't produced documents that don't discuss future plans. We certainly have --

THE COURT: But what I'm hearing from Mr. Crosby's argument is that these are, for example, documents, and the ones that I saw which you say are cherrypicked, but I have to say, looking at them, I'm not sure that that designation should apply or that it should apply the way that we're talking about or that it doesn't fall under the "or other person who otherwise possessed or knew the information" catchall, OK, the protective order.

If you want me to go through and rule on your various hot docs, I can do that, but it's going to cost you. It's

going to slow everything down. OK. We're talking -- you're talking at opposite ends. You're talking about documents for which there is no foundation, right. No foundation to show the witness.

That's not what I'm hearing from Mr. Crosby, that there is some foundation to show the witness, granted you may surprise them with it because they didn't know that the document existed, but that doesn't mean that they didn't have any otherwise possess or know the information or were aware of the communication or the meeting or the business development that was being discussed in that document, right.

So where do we find that middle ground?

MR. HURST: Your Honor, this is why the provisions in the protective order that allow for challenging designations or the procedure that we propose whereby they can show us documents in advance are adequate to the task without a wholesale modification.

If Mr. Crosby wants to show us documents in advance under the condition that we not disclose them to OpenAI, we are perfectly willing to do that. The courts have adopted that procedure, the procedure requiring advance notice in other cases. I have tried many cases where cross-examination documents were required to be disclosed 24 hours in advance of the examination with a carve-out only for true impeachment, your Honor.

And if it was true impeachment, that can only happen if the witness has seen the document before, so...

THE COURT: What about refreshing recollection?

MR. HURST: Your Honor, I don't see why that can't be disclosed in advance to us, under an undertaking not to disclose it to the witness or to counsel, and they always have the opportunity to challenge, you know, the designation of any particular document.

THE COURT: Right.

What I'm hearing from the plaintiffs here is that they are challenging -- you're challenging the designation on grounds that if 60 percent of the documents are being marked AEO, they can't all be AEO, is that what I'm hearing.

MR. HURST: Your Honor, just to be fair, we have resolved that issue. We had an agreement of counsel, we adopted new protocols, we completely reviewed our productions. And, your Honor, counsel, if that is the issue, he is reneging on an agreement here.

And I just think that's not appropriate, your Honor. I really --

We have resolved that issue. We undertook a massive effort in light of our resolution of that issue, and there was never any discussion at that time that the deposition carve-out would involve showing our documents to OpenAI. Here are all the things we agreed to. We agreed they could show any

Microsoft document to any Microsoft witness whether or not they had seen it before or there was any reason to think that they had. We agreed they could show Microsoft documents to a limited number of in-house New York Times people.

We agreed to all of those things because the only thing that was ever discussed at the time of the prior AEO dispute was that we would need a deposition carve-out for Microsoft people. This show Microsoft documents to OpenAI people thing came completely out of left field at the last minute prior to the court's deadline for filings. It was not something that was ever raised before and it was not part of any agreement.

We resolved the AEO issue, your Honor.

THE COURT: All right. Let me hear from Mr. Crosby.

MR. CROSBY: That is absolutely untrue. We did not resolve the AEO issue. We agreed to leave the deposition issue for another day because the depositions were not happening yet.

THE COURT: Right.

MR. CROSBY: So we had immediate need in order to advise our client to be able to show them these over-designated documents, the most important documents in the case, so they could assess how the case is going.

We made an agreement that we would be able to do that with respect to a small number of The New York Times insiders, but we specifically reserved this issue. And we have made

progress on this issue. Don't get me wrong. The percentage of designations is down from 98 percent to 60 percent. We do appreciate that. We don't think it's enough.

We don't think it solves the problem. We did agree, they did agree, and we appreciate it, that we can show documents to Microsoft witnesses even if they aren't the author of the document. But what we don't have, and where we're not yet, is these documents that clearly they document --

Outside of this courtroom, Microsoft and OpenAI are partners. This Oracle deal that was just announced yesterday, Microsoft is one of the technology partners for this. Partners, we all remember from law school, are honest with each other and don't keep secrets from each other.

But inside this courtroom, they are competitors. And all of a sudden, Microsoft's internal documents saying this is what we did with OpenAI are somehow supposed to be hidden from OpenAI. The notion that I want to go into court and use documents for which there is no foundation to examine a witness.

First of all, what would be the point of doing that? Second, all we're asking to do is to treat the documents that they've treated, that they have designated as highly confidential, under the same language which is in the protective order with respect to confidential documents. And that language says a witness, it needs to be reasonably

necessary, and the witness has to do a confidentiality undertaking. And the protective order that we have agreed to it in this case contemplates witnesses and experts are going to do a confidentiality undertaking without first seeing all the things that they are supposed to keep confidential.

So the notion that you couldn't possibly do that without seeing the documents first doesn't make any sense to me at all. But, look, your Honor, I mean, the bottom line, right, I think your clarification that you view this knowledge of the information like I do broadly and that it would be abusive for someone in a deposition to be nitpicking that in a constructive way, I think we might -- we could live with that clarification.

We certainly -- our preference would be, because most documents should be, at most, confidential, that we should treat all these documents confidential unless they are willing to bring the percentage down lower. But if that is not what the court's ready to do today, I think we could maybe take some depositions with the clarification that you have given us today and with the caution that you have given today and see how it goes.

THE COURT: I mean, look, I saw four documents, right. And even assuming they are cherrypicked to be the most obvious cases, they may be internal documents, internal Microsoft documents, that are talking about collaborative work or meetings happening between Microsoft and OpenAI.

I don't know how, if we're talking about those particular documents, why they couldn't be shown at a deposition with an OpenAI witness.

I mean, am I hearing any argument about that?

MR. HURST: Your Honor, I think that the examination of the documents in advance on a document-by-document basis would afford us the opportunity to have that discussion. Certainly there are documents where people privately express thoughts that were never meant to be shared with another person and, frankly, the only reason to show that to that person is with an objective to create or sew discord.

And I don't want to us see getting into that situation, and I think that that is an appropriate situation for objecting to showing such a document. You can always ask people about facts. You can always ask people about the conversations that they have had and the underlying facts. There is no concern whatsoever about that.

THE COURT: OK.

MR. CROSBY: Asking someone about the underlining facts when you have a document documenting the facts and you can't show them, often results in a person --

THE COURT: This goes back to my question what about documents to refresh recollection, right?

MR. HURST: Right.

THE COURT: You were talking about impeachment

documents. And with impeachment documents, a recipient had to have seen it. But refreshing recollection is a different animal.

So how do you propose to deal with that?

MR. HURST: If on the face of the document it describes conversations between the person who is in the chair and others, I agree that that is something that is appropriately shown to that witness under the current and existing terms of the protective order without need for modification.

THE COURT: All right.

MR. CROSBY: I do want to come back to this one example, you know, that if we can't deliver on this, OpenAI will be very disappointed. I think that strongly suggests that there may have been some subsequent interactions or lack of subsequent interactions and that, sort of, I think that should also be fair game.

But this is litigation on people. It's not about hurting people's feelings. The reason that things should be highly confidential is because they are competitively sensitive, that OpenAI would gain a commercial advantage against Microsoft by knowing this. That Microsoft -- that OpenAI's feeling might be hurt by knowing something Microsoft said is not a reason for something to be highly confidential.

THE COURT: I tend to agree.

Look, I'm not going to tell you how to conduct your deposition. You all know how to work around these things. I don't expect to see motions made under Rule 30 where I get the entire deposition transcript and you're pointing out, you know, the number of times that there have been objections made or something.

You're all beyond that. You're all above that. Everybody, even the lawyers not speaking today, have probably gotten more depositions under their belt than I did back then. OK. You will find a way around this. You will find a way to work with it the way that it is. I'm not directing anybody to go re-review and re-designate.

I think given what we talked about today and these particular documents, I think there is a fairly wide area now where we have some clarity about where you can work, where you should work, you know, showing, you know, presenting some documents 24 hours in advance isn't necessarily a terrible thing, but I'm not going to order that it happen either.

You're all far more experienced in taking and defending depositions. You know how this works. And if there is a motion under Rule 30, we'll deal with it then. But I stand with what I said earlier about, you know, I expect that objections should be made in good faith. Even in the depositions, you're trying to work all of these issues out in good faith. Make the record on the, you know, in the

deposition transcript.

If you need more time because of it, you might get more time. All right. We're not going to be rigid about this, but at the same time, I really expect you all to try to work through this.

Like I said, the examples that I saw, admittedly, are a tiny, tiny fraction of what has been marked AEO. If we start to get through the edges and you're really finding yourself at an impasse, we can revisit this. I really want to encourage you all to really to try to work through this and talk through this, given how we have discussed it today.

MR. CROSBY: Just so that I don't get in trouble down the road.

So, for example, if I were to use the four documents that we submitted, illustrations in support of this motion, with an OpenAI witness to question them about this interaction with Microsoft, I would not -- in your view, I would be within the bounds of the protective order as it exists because that person would have knowledge of the underlying circumstances.

Am I understanding correctly, your Honor?

THE COURT: It sounds like that is what Ms. Hurst was saying, also.

MR. CROSBY: I hope so.

THE COURT: You know, I don't ...

Yes. I mean, nevermind.

MR. CROSBY: Those would be fair game. Those are examples of what would be fair game documents.

THE COURT: Yeah. I think what I'm saying is, I have seen in other cases motions, or gotten calls during depositions -- which I really, really don't think should happen here -- but I've gotten calls in depositions where the attorneys are fighting about, you know, he's objecting too much, he's making speaking objections, da, da, da.

I don't expect that to happen in this case, OK. And this is the type of issue where I'm not going to give you the confirmation that, like, you can use, you know, document 386-2 -- I don't even remember what it said -- on every single OpenAI witness, right.

Obviously you're going to make your determinations about whether or when to you use it, whether and when it might be appropriate. And it may depend on who is copied or, you know, on the document and how the deposition progresses. I'm not going to tell you how to do your deposition. But we have Rule 30, we have all of your experience navigating around these very sensitive issues.

And, you know, most of the deposition, it seems, is going to be about issues and times where Microsoft and OpenAI were collaborating or working together on projects. So there are a lot, there is a lot of ways to work through this, right.

MR. CROSBY: OK. So, again, if I've got a Microsoft

document that talks about an interaction with OpenAI and I'm not going to get whacked or, at least, you know, for using it. Obviously the circumstances will be the circumstances.

THE COURT: Yes.

MR. CROSBY: But there is no hard-and-fast rule that says you can't use a Microsoft document that OpenAI didn't see to examine an OpenAI witness when the document talks about an interaction between or suggest an interaction between OpenAI and Microsoft, am I correct?

THE COURT: That seems to me to be the case. There is no hard-and-fast rules -- I'm not giving you an advisory ruling.

MR. CROSBY: OK.

THE COURT: -- on documents I haven't seen. I don't know what you're doing with the depositions.

MR. CROSBY: I will stop seeking further clarification, and I will take the words that the court has expressed and, hopefully, so will my opposing counsel.

THE COURT: Again and again, I mean, I do -- I'm not averse to the sharing some documents 24 hours in advance. OK. I mean, 24 hours in advance, when you're prepping for some of these types of depositions, you're still going to get your element of surprise.

And, by the way, like, the aha moment is much better on the witness stand than in the deposition transcript.

MR. CROSBY: True. True. But Steve Susman would still be rolling over in his trial boots.

Thank you, your Honor.

THE COURT: I get it.

MR. CROSBY: Thank you, your Honor.

THE COURT: All right. I thought that was going to be a quick issue.

All right. The searches on AEO training data. I think that is still premature, right? You're still meeting and conferring?

MS. MAISEL: Your Honor, if I may.

THE COURT: Oh, no. OK.

MS. MAISEL: Jennifer Maisel from Rothwell Figg on behalf of The Times and Daily News plaintiffs.

Your Honor, we're making some progress. As you may recall from our last conference, we discussed a two-stage approach. There is this foundational issue of what works were used to train the GPT models at issue.

We landed on, first, the identification phase of that approach and two searches, a URL keyword search and then an Ngram search. And we made substantial progress on the URL search, as in the joint chart, OpenAI is committing to producing those search results before the end of the month.

In contrast on the Ngram search, we are still meeting and conferring about that. There's been significant delay. We

initially proposed an Ngram search back in November, two months ago. After meeting and conferring, OpenAI rejected that Ngram proposal. We asked them for a counterproposal. It wasn't until the day before we filed this joint chart that we finally got a counterproposal.

So I would just submit that we work well with deadlines, and I would ask that the court order us to provide a status update in the next week or so. And as part of that update, if we cannot come to an agreement on the Ngram search, I would ask that we -- I would ask for a quick ruling on that, just because the searches take quite a while to complete and we would like to move this aspect of discovery forward.

THE COURT: All right. In my notes I had said, although it's premature, please still meet-and-confer. I had in my notes, Encourage OpenAI to keep plaintiffs apprised of goal posts for timing.

All right. So do you want some timing, some timing goal posts?

MR. BAILEY: Your Honor, if I may. Ed Bailey for OpenAI.

So yes, everything Ms. Maisel said is correct. We are continuing to meet-and-confer on this issue. On the Ngram search specifically, we are still waiting on some information from The New York Times about the scope of what that search might entail. This is something we asked for several times,

and there is clearly some miscommunications, which is apparent in the chart.

But what we need from them -- and this is really going to be necessary for us to assess the burden and scope of this search -- is identification of the specific datasets that they need to search. Because as part of this Ngram search, we have to set up a customized environment within OpenAI. We have OpenAI engineers working on this issue. It takes engineering time for every additional dataset that they need to search. A lot of these datasets are not going to be relevant to finding The New York Times articles and New York Daily News articles.

So we are trying to work with them on how can we limit the number of relevant datasets so we don't have to waste our time setting up all of this infrastructure for datasets that aren't very likely to have any information. So that is, sort of, key issue number one.

The second issue is the size of the Ngram. The N in the Ngram, right. When we last met, my colleague talked about what a six Ngram might be, six words in a row. That is going to obviously lead to a lot of false hits. There is a lot of six words that show up, not just in The New York Times articles, but in things all over the internet and everywhere.

We have been trying to negotiate with them on what is a reasonable N, what is a reasonable set. We think it is one or two paragraphs. They are pushing back on that. We need

what their position is going to be so we can, you know, come to a head on this in terms of whether or not we can agree with the Ngram search.

So, if your Honor is going to put a deadline in, what I would ask is to get a deadline first from them, when they are going to provide the positions on these two questions, and then we can have a chance to respond and finish this meet-and-confer.

THE COURT: Ms. Maisel.

Speak from there if you speak up.

MS. MAISEL: We can commit to providing the datasets to search by the end of this week. And in terms of the threshold, I think we settled on the Ngrams. I think there is a different threshold issue. But that, as well, we can get you by the end of the week.

MR. BAILEY: That would be helpful, your Honor.

THE COURT: OK. So that would be the 23rd. The New York Times to provide datasets for the Ngrams.

OK. And then should we have another back- and-forth? I mean, I hate to micromanage your timing of your meet-and-confer, but if they get you the information by the end of the week, what's the next step and how much time does that take?

MR. BAILEY: So, your Honor, we would -- we would need some time to work with the engineers. Obviously, I think, having a week to do the meet-and-confer and then be able to

work with the engineers. I think two weeks would be ideal because that would give us the time to both meet-and-confer with the New York Times plaintiffs and also work with the engineers before we, sort of, have it finally teed up.

I don't think we would need more than that.

THE COURT: OK. So that would be, what, February 6?

MR. BAILEY: Yes.

THE COURT: OK. So tentatively we will aim for a status letter February 13.

MR. BAILEY: That works for OpenAI.

Thank you, your Honor.

THE COURT: OK. All right. This was a short issue. The next one might be longer. And they are, it starts with ECF 379. I believe it's output log data or ChatGPT user conversations. You might need to bring me up to speed a little bit on what those are, and then I see as possibly a related issue is the Microsoft data logs.

Are those kind of the same?

MS. MAISEL: They are related discovery issues, but separate legal issues, I would say.

THE COURT: OK. Let's start with the output log data.

MS. MAISEL: OK. On OpenAI's output log data, this is ECF 379.

This is Jennifer Maisel again.

As outlined in our letter, it's recently come to our

attention that OpenAI has not been preserving all user conversations, and these user conversations are the prompts and outputs that end users are providing to their products, including ChatGPT.

THE COURT: This is all OpenAI users?

MS. MAISEL: Yes.

THE COURT: OK.

MS. MAISEL: And we were informed by a letter dated November 15 that in accordance with OpenAI's website policies and contracts with its end users, it had been deleting some set of this user conversation data for some unspecified time period. And we responded back asking them to identify what has been deleted, whether any of that deleted data included news plaintiffs' content, referenced news plaintiffs, and whether it's recoverable.

We did not get an answer. We also asked OpenAI to suspend its deletion policies and to preserve data containing news plaintiffs content on a going-forward basis. Again, OpenAI is not committed to do that, and the relief that we're asking for right now is we need OpenAI to identify with specificity what has been deleted, what volume of these user conversations have been deleted over what time periods, was this data deleted, for which products.

We have a couple products in the case. It's ChatGPT, the browser plug-in, search GPT, the API, and certain custom

GPTs, and whether any of this deleted information is recoverable. And most importantly of this deleted information, did any of it contain news plaintiffs copyrighted works, whether through regurgitation issues with the models or due to retrieval augmented generation, where OpenAI's products is conferring with Microsoft's Bing index to retrieve news plaintiffs' content from their websites. Right now we're asking for a specific identification of what has been deleted as well as for OpenAI to preserve this information moving forward.

There is one point I want to highlight for your Honor with the relevance of these logs and why they are so important, especially the earlier ones. Just last week we had a motion to dismiss hearing before Judge Stein. And both defendants have moved to dismiss the contributory infringement claims based on an alternative theory of ours that, to the extent the end users are direct infringers, that end users are using ChatGPT and the other generative AI products to infringe the plaintiffs' copyrighted works.

We want information about that end user activity. OpenAI has represented that this is a fringe behavior. No normal user is using the products to bypass pain walls and to retrieve copyrighted content. But all of this information is going to be in the logs.

And as we have learned through the course of

discovery, we are undertaking source code review, we have been evaluating the training datasets, and it has become clear that OpenAI has taken intentional steps to make it more difficult to collect this evidence on a going-forward basis.

We are in open court, so I'm going to be very vague here. But there are specific logs on these specific plaintiffs that have sued them in litigation. Not for all copyright owners generally, but for the specific plaintiffs, and it is an evidence issue, a burden issue that plaintiffs bear. This evidence is uniquely in the possession or was in the possession of OpenAI, which is why we're asking for them to identify what was deleted and when.

THE COURT: OK. When we had our last conference, we touched about this issue because it was relatively new at the time. And I thought that you were still meeting-and-conferring on trying to get this information back then. I'm just trying to find out where we were when this was raised last time.

MS. MAISEL: So, at our last hearing, we had -- we were seeking an early custodial 30(b)(6) deposition of Microsoft to ask information of Microsoft about its output logs. And pursuant to your court's guidance, we engaged in a more important informal meet-and-confer process. We were working on setting up a meeting amongst the technical experts to evaluate our query to the datasets.

We did not talk about OpenAI's open log issues, but we

could have a 30(b)(6) custodial deposition of OpenAI that we are in the process of scheduling. It's scheduled for next week. The parties have been meeting-and-conferring about the topics for that deposition. One of the topics we have asked for relates to custodial issues around the logs.

And as of our last meet-and-confer, OpenAI was objecting to putting up a 30(b)(6) custodial witness on that issue, on that topic, so that it would need to be a different deponent than the other topics, and suggested that we pursue a more informal process.

But I would submit, your Honor, that we really need a statement under oath or a representation again about what OpenAI's custodial practices have been with respect to these logs. Again, identification of what has been deleted, including information specifically relevant to these cases.

THE COURT: OK. All right. Before I hear from OpenAI, what you've told me so far is that the news plaintiffs are seeking the lookback, right?

What's been destroyed or not preserved, what did it contain looking back?

Are you also asking for preservation going forward and is that also a dispute?

MS. MAISEL: Yes.

We have asked for preservation moving forward. We asked for them to specifically preserve evidence relevant to

the litigation. They have come back and said, we don't have a solution in place to identify --

THE COURT: In other words, yes, how do you identify which output logs are relevant.

MS. MAISEL: Exactly.

THE COURT: OK.

MS. MAISEL: So our position is, if we can't do that, you know, for the time being, our position would be preserve everything and we can sort this out based on that preservation.

I will add, in terms of a burden issue, what is missing from OpenAI's response is there is no articulation about a specific burden or cost associated with the storage of this data, you know, what volume of storage that might be.

And I would note in their online -- in their online privacy policies and their online terms, they do have exceptions to preserve this data for legal purposes. And in their opposition, they raised issues about reputational issues, their obligations under privacy laws.

We are more than willing to work with OpenAI to preserve the privacy of its end user. We are not seeking information about user's personal identifiable information. Again, our discovery is tailored to the copyright issues and other issues in these cases.

And just as another point of contrast, Microsoft is preserving all its data. They have represented to us that they

have been preserving the similar user conversation or chat log data since the inception of the litigation, so just another point of contrast on burden.

THE COURT: OK. All right. Let's hear from OpenAI.

MS. NIGHTINGALE DAWSON: Good morning, your Honor. Elana Nightingale Dawson on behalf of OpenAI.

I was heartened to hear Ms. Maisel say they are willing to work with OpenAI with respect to concerns about privacy. I will start by saying that what plaintiffs are asking with respect to preservation of all user conversations on a going-forward basis is a request to disregard the expressed preferences of OpenAI's users.

And to be clear, the situation is different in kind from the situations in the cases that plaintiffs cited and what plaintiffs are describing. The suggestion that data does not exist for the lookback period that they referred to.

Without getting -- I'm going to speak in generalities because the numbers are under seal. But at Docket No. 425, that is the declaration from Michael Trinh, at paragraph four, it details the many, many billions of conversations that are retained. And that is because OpenAI's default practice, its default policy is to retain all conversations unless the user ChatGPT user has an expressed preference to not have the conversations retained.

And so it is as much in OpenAI's interest as it is, as

plaintiffs desire, to have these conversations exist and that is why so many billions exist. We are already at a place that there are so many conversations that the parties recognize that it is not possible to actually access and work with all of that data. And that is why we are engaged in conversations about sampling that data to come up with a representative sample from the time period in the past from the billions of conversations that are already retained in this situation.

Plaintiffs are now asking not just that the conversations that relate to them be retained, but that all preferences from OpenAI's users be disregarded on a going-forward basis.

THE COURT: Let me stop you right there.

So I'm not suggesting that this has happened, right. But hypothetically, say, a ChatGPT user who had been using, found some way to get around the pay wall, right, and was getting The New York Times content somehow as the output, found a way to do it. And then hears about this case and says, Oh, woah, you know, I'm going to ask them to delete all of my searches and not retain any of my searches going forward.

I mean, isn't that directly the problem?

MS. NIGHTINGALE DAWSON: So I don't think there is any reason to believe that that -- to be clear, I hear what you're saying, your Honor. I think what we already have -- again, I don't want to say the specific number -- but the volume that we

already have, the billions upon billions of conversations that are already retained, to then drive a truck through the retention for legal purposes that Ms. Maisel referred to and say we're actually going to disregard the expressed wishes of OpenAI's users who request not to have their data preserved, on the off chance that that has happened, I think is really putting the cart before the horse because of how much we already have.

And we are happy to -- I do think your Honor has said in a number of circumstances, it sounds like the parties are still conferring. I will note that what the plaintiffs originally asked for was that we look at ways of addressing their content in particular, and we have been actively working on that. And so this is a situation where we would invite the court's encouragement for the parties to continue discussing.

We certainly hear what the plaintiffs are saying, but as your Honor noted and as Ms. Maisel said, this is also a matter that is related to what plaintiffs are already engaged with with Microsoft, the informal technical discussions. And so we do believe that rather than issuing a carte blanche blanket order for retention, that we should continue to engage in these discussions because, again, we have so much data already that we want to be able to find a way to make that accessible and usable. And it does cover this entire time period. And, frankly, goes from before this lawsuit, the

lawsuit was filed. And so we're not just talking about a finite period of time, we already have voluminous amounts of data for this time period.

THE COURT: OK. Let's talk about the going-forward preservation.

Is there a way, is there a way to segregate the data for the users that have expressly asked for their chat logs to be deleted, or is there a way to anonymize in such a way that their privacy concerns are addressed or, like, what's the legal issue here about why you can't, as opposed to why you would not?

MS. NIGHTINGALE DAWSON: Sure.

So I want to make sure I'm understanding. I heard you say the legal issue. I'll start trying to answer your Honor's question. If I'm not getting to your point, please let me know.

So, at docket 425, this is the under-seal declaration submitted by Mr. Trinh, we detail in perhaps eight, nine -- eight and nine, the specific issues. And so really, I think, the details are what is at issue here.

So what I heard you saying, your Honor, was asking about what the user has requested. What I understood plaintiffs to be asking for is they are focused on what comes out, not necessarily what the user requested. And we detailed here -- I mean, this somewhat relates to the training data

conversations we have been having and otherwise -- of exactly how you undertake that type of analysis, how do you even begin to identify what that is.

So, we are happy, I will say, that on, I believe it was, January 9, Ms. Maisel mentioned a few other technical functions and filters that they wanted us to discuss the possibilities of using. Based on what they their ask is and the technological measures they are pointing to, there is a mismatch. But I do think we can continue to discuss with them exactly what they are trying to look for and if there are ways of trying to address it.

But I would say the, kind of, carte blanche, preserve everything request is problematic, as is, I think, what your Honor said for the reasons set forth in the declaration, the more specific targeted piece.

But, again, to be very clear, we are retaining and that is our default position, and we are continuing to retain as the default position.

THE COURT: OK. So talk to me about --

I think I'm, like, way off on the side in the fringes of the problem. So when you're talking about -- you are currently retaining, the default is retaining the logs, except for users who have specifically said delete my logs.

Is that a going-forward ongoing basis, or is it that a user can say, you know what, maybe you were retaining before,

now I want you to go back and delete everything that I have engaged with the GPT?

MS. NIGHTINGALE DAWSON: So there are different -- there are different options. All of those are possibilities.

I should say that the default we were talking about for ChatGPT and then the API, there are specific policies there. But with respect to, a user can designate a conversation as temporary -- this is paragraph six of the declaration at docket 425 -- a user can elect to delete a specific ChatGPT conversation or they can make a request to delete an entire conversation.

I should say, there are numerous privacy laws and regulations throughout the country and the world that also contemplate these type of deletion requests or that users have these types of abilities. So it is a combination of our commitment.

I mean, as I think we see in the news and otherwise, users of technological products are very concerned about having the ability to control what they are doing in the data. This also interfaces with the commitment to privacy as well as your obligations under various laws.

THE COURT: All right. So I think I'm seeing a little bit of the mismatch of the ask, which is plaintiffs are asking for the conversations to be -- everything to be preserved going forward, right.

But how does this work with the interaction with the users? Right.

Because if the users are potentially in realtime, I mean, is this a tech -- is this a technological problem also, or is this, like, how do we even get through this then?

Because I think I see what you're saying, is that you could have a user who is, right now, right, doing things and then going back and deleting or deleting their own chats. No?

MS. NIGHTINGALE DAWSON: Yes, I think you're exactly right. There is a technological issue in that -- I do think this kind of reinforces why it would make sense at this juncture to require the parties to continue conferring on this issue.

I understand, I mean, one of the reasons we wanted to wait until we actually had information was because the ask is so significant. The specific ask came in before the holidays. It took us some time to figure out even what would or wouldn't be doable from an engineering perspective and that is what is set forth in the declaration.

But I do think, similar to the conversations that are going on around outputs with respect to sampling and how we undertake that, I do think further conversations between the parties would help us continue to move forward. Because of as your Honor said, these technological issues and, frankly, the user privacy concerns that, as I said, I was heartened to hear

plaintiffs say they are mindful of.

THE COURT: OK. Go ahead, Ms. Maisel.

MS. MAISEL: There is just a couple points I want to address.

First off, I think your Honor hit the nail on the head. The users that are likely to be deleting their conversations are potentially the ones using these tools for wrongdoing.

Imagine a user who had been using ChatGPT's remove pay wall custom GPT to do exactly what the name suggests, and they see the lawsuit is filed and immediately ask OpenAI to delete their logs. And I think this is precisely the problem with the retroactive lookback.

We don't know what was deleted. We don't know if any of those logs contained information that's relevant to this case. We don't know if OpenAI can answers those questions at this point right now.

THE COURT: OK.

MS. MAISEL: And prospectively moving forward, the fact that we can't segregate this data out. And, again, OpenAI has not articulated what burden there is in preserving everything.

THE COURT: Well, I think that is the burden. I mean, the burden isn't just the storage, right. It's the competing needs and trying to address the requests, the privacy concerns,

and how --

I mean, there is a technological and engineering burden, I guess is what I'm saying, as well. Because how do you manage this and not lose anything or do you just override what people want to delete?

There could be also possibly -- I mean, I've never used ChatGPT. This is all very hypothetical for me. There could also be people who are, like, write me something, you know, smutty and then they are embarrassed, right, and they want to delete it. There are so many other reasons that somebody might want to delete their chat. This is also why I've never done it, but I don't even want to go into that area.

But it's -- but I think one of the concerns is, I saw the burden as potentially being storage. But also what is coming clear to me is that there is a technological issue about how do you manage or how do you thread this needle with, you know, what might even be happening realtime with the -- I don't even know if it's millions or billions of users of ChatGPT. Even as we speak, right, there are people who are possibly doing this realtime or have things set, you know, have their preferences set in such a way.

Can I ask you to keep talking about this and particularly about the technological and engineering ways to manage this issue?

And then if you want me to give you some timelines to

talk about it so it's not -- you're not having the conversation the day before a status letter is due, we could maybe get a status on this on February 13, also.

MS. NIGHTINGALE DAWSON: That would work for OpenAI, your Honor.

MS. MAISEL: That works for us, your Honor. But it's only part of our request --

THE COURT: OK.

MS. MAISEL: -- under this order.

The other part of our request is the specific identification of what has been deleted, whether it's recoverable, what volume of data it is, whether it's relevant to issues in the case, whether it contained news plaintiffs' content or other intellectual property.

We do need answers to these questions.

THE COURT: OK.

MS. NIGHTINGALE DAWSON: We are certainly happy to discuss with Ms. Maisel.

I will note, it's a bit surprising to get this now. I did hear Ms. Maisel say that she did not know about this issue until November. That is surprising because our letter from February 29 -- and this is at our letter from February 29, which was attached as an exhibit, this is at docket 379-3 -- went to Mr. Lieberman and to Ms. Maisel at Rothwell Figg, where we explained we were supplying them with our privacy policies

and what that meant.

So, frankly, I don't know that what plaintiffs are asking for is possible, but happy to continue discussing with them the concerns that they have. And I think I would ask, from our perspective, let us discuss that part of the request as well as an update, your Honor, on the 13th of February.

MS. MAISEL: I'll just respond to that briefly.

THE COURT: OK.

MS. MAISEL: ECF 379-4, this is Exhibit 3 to our motion, there is an e-mail dated March 5 from Ian Crosby saying, We disagree with respect to OpenAI's position on its preservation obligations. We will comply with our preservation obligations --

THE COURT: Wait. This is 379-4?

MS. MAISEL: 4.

THE COURT: OK.

MS. MAISEL: Page two.

THE COURT: OK.

MS. MAISEL: And this was the last word on the issue. We had the Rule 26 report. We had many requests for production, requests for inspection, and not a single one of OpenAI's responses and objections to our discovery requests did they say, we are only preserving these categories of output logs.

To be clear, when I say output logs, we are looking

for the prompt, and I think that's clear from our discovery request, and not once did they raise this issue.

Keep in mind, the Daily News complaint was filed after this February date. The Daily News complaint wasn't filed until April, and it wasn't until mid November that we received a letter that said, We wish to remind you of our preservation practices.

And the Daily News case, there never was such a letter. There was no reminder of anything. And in The New York Times case, we thought we had resolved this. This issue had not been brought up again. We had countless meet-and-confers about our discovery requests. OpenAI did not move for a protective order under Rule 26(c).

THE COURT: OK.

MS. MAISEL: So we were, quite frankly, surprised to get this letter.

MS. NIGHTINGALE DAWSON: May I just respond on one point, your Honor. This may be in the weeds a little bit.

I would just like to point out that the e-mail that Ms. Maisel pointed to at 379-4.

THE COURT: I read it. I just looked at it.

MS. NIGHTINGALE DAWSON: OK. So what I would like to point out is, at the bottom of that, it is a March 5 e-mail that responds to and it was a letter regarding plaintiffs creation of Exhibit J. So that letter was not submitted on the

record and I have one copy of it.

We could get it to you. But my point is simply that the e-mail from Mr. Crosby where we were saying we disagree was referring to a letter that is not in the record. It was not in response to the February 29 letter. But I really do think this is ultimately all matters that the parties can discuss and try to address given the competing issues that are being raised and certainly understand.

But, again, the overriding privacy concerns that we have as well.

THE COURT: Look, I also think at some point we are where we are. We now know that this is what was or wasn't preserved. We are getting a better understanding of the technological issues maybe of trying to find out what's been lost, and I would like us to try to address the issue at hand without getting into a spoliation argument right now.

OK. Let's see what we can -- let's see what can be figured out, what can be ascertained. Let's see what, if there is a way to find out what was deleted or not preserved in the past. But then also, going forward, if there is a way to manage what seems to be fairly clear-cut and specific concerns that may not -- that may only be a small subset of the users and the logs that want to be deleted, right. See if there is a path going forward where you can manage some of that.

OK. Microsoft data logs.

MS. NIGHTINGALE DAWSON: Thank you, your Honor.

MS. MAISEL: That's me again.

THE COURT: I figured.

MS. MAISEL: I'll stay here.

So we have made progress with Microsoft on the output log issue. We have exchanged letter correspondence.

I would say, again, this might be the theme of these inspection requests. Having a deadline helps the parties move this process forward. Plus, plaintiffs sent a letter with informal questions and technical issues that we would like to be addressed on this forthcoming call between Microsoft's technical experts and witnesses and news plaintiffs' technical experts.

We have received Microsoft's response to that letter. Again, the day before the chart was due, after we sent our position statement saying we had not received a response. And I think if we can get a deadline on the calendar for that technical conversation to occur, that would be immensely helpful. I think we have a lot of helpful information from Microsoft so far, and we feel ready to move forward with that conversation.

THE COURT: OK. February 6 to have the conversation, the technical conversation, by February 6.

MR. HURST: Your Honor, I have to consider the possible impact of the Chinese new year on the availability of

people on our end who might need to participate in this. I don't have those dates memorized. I apologize.

THE COURT: I think it's January 28.

MR. HURST: It's a whole big deal for some people --

THE COURT: I know.

MR. HURST: -- who are relevant to us here.

Your Honor, I think I might need a little leeway for --

THE COURT: Let's say for February 10 then.

How is that?

MR. HURST: Yes. Thank you, your Honor.

THE COURT: OK.

MS. MAISEL: Thank you, your Honor.

THE COURT: OK. February 10.

MR. HURST: Your Honor, may I make one followup request related to that?

THE COURT: Sure.

MR. HURST: We did produce quite a bit of additional information, including multiple new samples, specifications, and answered questions. If the court could ask the plaintiffs provide us with any additional questions that they expect to discuss on that call occurring prior to the 10th, if we could have that at least three business days in advance of the scheduled call, that would help us prepare.

THE COURT: Maybe February 6, keeping in mind the

Lunar New Year issues.

MR. HURST: Thank you, your Honor.

THE COURT: OK. They should have a week to recover from Lunar New Year.

OK. Custodians issues. There were the four that we discussed in December, which I think are still not ripe because you haven't gotten the documents. I'm trying to remember whose custodians they were.

MS. SOLOMON: Your Honor, this is Sarah Solomon on behalf of OpenAI.

Since we were last before you, The Times has produced a total of 355 documents --

THE COURT: Stop. We are going to talk about custodians right now. I want to talk about the two new custodians who are apparently pay wall or access controls.

Why are these -- what do you need to know, how is this relevant to claims and defenses, and how is this different from custodians that are already being produced?

MS. SOLOMON: Sure.

THE COURT: Why these two new custodians?

MS. SOLOMON: Sure, your Honor.

These are two custodians, two engineers at The Times that have knowledge of The Times' use of access controls and the way that their pay wall might interact with bots or crawlers. And The Times' knowledge of the extent to which and

the manner in which its asserted works are cached or indexed by such crawlers is relevant to potential defenses of waiver and implied license.

And the issue here, The Times, we have agreed on search terms related to these subjects, and The Times agrees that they are relevant subjects. The problem is that they are running those search terms against quite high-level employees. And so it's, like, their CTO or their head of product that are just not likely to be the ones that are having these types of conversations about technical capabilities.

They have identified these two individuals specifically in rog responses as knowledgeable, and so all we ask is that they are added and, at that, only for these specific search terms.

THE COURT: OK. Let's hear from The New York Times.

MS. PEASLEE: Good morning, your Honor. Katherine Peaslee for The New York Times.

So we disagree with the notion that these two individuals are the ones most likely to have documents on this subject and, therefore, the problem is that The Times has just designated the wrong custodians. We have designated Christine Liang, who is the senior director of technical search. I think of all people, she is the most likely to have documents on this topic, to the extent that you're going to find the information that defendants are looking for in The Times' e-mails.

But I think a more important point is that the sort of information they are looking for -- for instance, when The Times exclusion protocol was in place -- is not information that you are going to most efficiently or most likely find in e-mails.

We think better sources of that information would be, for instance, The Times' Jira database.

THE COURT: Which database?

MS. PEASLEE: Jira. It's a software program that The Times engineering and products teams used to track various issues. They submit issues to that database or to that system. That software they turn into tickets. Those are resolved.

If you have ever submitted something to IT and received a ticket back, it's probably using a similar sort of software. We have provided all of the Jira tickets back to 2012, which is what The Times has available that hit on the robots.txt.

And we would be willing to talk to defendants, if there are other, sort of, targeted searches they would like us to run to try and get at these topics. But The Times is also continuing to produce documents from, for instance, Christine Liang, who we think is the better custodian on the subject. Again, to the extent that the custodian is the appropriate source for this information anyway.

I want to talk on the point Ms. Solomon made about the

fact that they are only requesting these custodians for a limited number of search terms. That is good. We appreciate that. But it overlooks the cost associated with adding a custodian in the first place.

In order to make somebody a custodian, The Times has to go collect all of their documents. Even to run search terms to get hit counts, The Times has to do that because, unfortunately, The Times is not available or able, due to the software that it uses, to run search terms in the native environment. We have to actually go collect all of those, deliver them to the document vendor, and have them do it. We have looked into switching that software, and, unfortunately, cannot do so on a timeline that is useful to this litigation. But we did see if that was an option.

And that process of collecting all of someone's documents is not just e-mail, although that is a big piece. There is e-mail/shared drives, personal drives, scanning the mobile device, scanning their hard drive, photocopying their hard copy documents, if they have those. All of that is part of that collection process.

And I reached out to our vendor to ask, what is the estimated cost per person just to complete on your end that collection, processing, loading, staging of data, to run search terms. And the answer I got was about $46,000 per person.

(Continued on next page)

MS. PEASLEE: (Continuing) And that is not an insignificant cost to incur just to get to that stage for custodians that The Times does not believe are relevant, important custodians.

And I am happy to answer any further questions that your Honor has about these particular individuals.

THE COURT: I'm guessing you've met and conferred and said exactly this, so I'd like to hear what OpenAI's response is to that.

MS. SALOMON: This is Sarah Salomon.

First, on the robots.txt files, it may well be that the appropriate repository for that is Jira. However, The Times has also agreed to run custodial searches, and should do so against the appropriate custodians. These would be the only two technical witnesses that The Times would designate as custodians.

But the more crucial point I want to address is Ms. Peaslee's allegations with respect to burden. And to borrow Mr. Crosby's phrasing, inside this courtroom, The Times has portrayed itself as a small company with a meager IT department, only one temp doing these collections, but to the outside world, it is a publicly traded company with significant resources, a repeat player in litigation, having been involved in over hundreds and hundreds of suits over the past ten years. It is not sufficient to say that it is unduly burdensome to

collect from two additional custodians, especially given the fact that these new facts that Ms. Peaslee has provided regarding their vendor were never before raised in a meet-and-confer, and have never been provided in a signed declaration to your Honor.

They are the plaintiffs here. They are the ones who brought this suit and should resource it accordingly.

MS. PEASLEE: Ms. Peaslee, on behalf of The New York Times.

To respond to the idea that The Times is not appropriately resourcing this case, I believe OpenAI themselves cited that The Times, in the first nine months of 2024 alone, spent $7.6 million litigating this case, and that doesn't even include the last three months of 2024, which your Honor is well aware were very busy in discovery.

So, it is not the case that The Times is not investing substantial resources in this case.

And we agree that we will provide relevant responsive custodians. It's simply that these two individuals they've identified are not that.

THE COURT: So, yes, let me hear from Ms. Salomon, because we're talking about burden, but before we can get to burden and whether it's proportional to run these search terms, what is it about their documents? Is it something that you've seen in the Jira tickets that suggest that these two

individuals should be additional custodians?

MS. SALOMON: Your Honor, it's actually something we have seen in The Times' interrogatory responses.

THE COURT: No, their names?

MS. SALOMON: Yes, they are named --

THE COURT: What else have you seen?

MS. SALOMON: Not much, your Honor, because The Times has not, outside of the Jira tickets, produced much in the way of custodial discovery about these subjects.

THE COURT: But I guess my question is, why is it that if you have specific technological questions, The Times didn't identify these individuals as working in this area, that you have the Jira tickets and you take depositions?

MS. SALOMON: Because The Times has also agreed to run custodial searches that we would need. We need those documents prior to the depositions to question witnesses.

THE COURT: I guess I'm still not seeing it if The Times is saying, look, they're not talking about this kind of stuff over email, right, this is not an email issue.

MS. SALOMON: But, your Honor, to be clear, my understanding is that that representation was specifically with respect to successive versions of the robots.txt file. I understood what Ms. Peaslee to be saying is the most efficient way to be collecting those successive robots.txt files is not email — fair enough. But that's not to say, and I don't

understand The Times to be saying, that its custodians are not having discussions about how its pay wall operates over email, or they are not having discussions with respect to, we are noticing the potential of a certain bot or company crawling our cite, what are we going to do about that. I've never understood The Times to be saying that. That is not something that would be included in custodial data.

THE COURT: Right, but isn't that something Ms. Liang would probably be aware of?

MS. SALOMON: That's something that The Times has represented they could substantiate that by offering hit counts showing that Ms. Liang is the one most likely to have the documents and not the two custodians that we've proposed and that they have identified specifically. The Times refuses to do that, again, based on unsubstantiated burden allegations.

THE COURT: This has a cost burden, and it is substantiated because it's $46,000 per person, right? So there is some substantiation.

The other thing is that -- have you gotten Ms. Liang's docs yet?

MS. SALOMON: No, your Honor. And that's what I was raising earlier. Since our last hearing, we've gotten 355 documents, total, from The Times. They have not produced from 10 of their 22 custodians, so we don't have custodial productions from near half of their custodians. And so when

we're having these conversations about, oh, you know, Ms. Liang is likely to have these documents, or someone else, it's all in the realm of the possible, but it's not based on documents that we've actually received.

THE COURT: Okay. But then the request to add additional custodians is also not based on what you've already received.

So, I'm going to deny this request without prejudice to raising after you get Liang's documents.

How is the production going? In other words, are we getting Ms. Liang's documents at the end? Are they coming?

MS. PEASLEE: Sure. Of course, your Honor. We currently have 110,000 documents from the ten newer custodians backed out for review, and we're adding to that population as we continue to come to ground on search terms and appropriate custodians for those search terms with OpenAI and Microsoft. We are happy to prioritize Ms. Liang's documents. We can put those at the top of the queue so that those are reviewed in the next couple of weeks.

THE COURT: Well, why don't you meet and confer about what you want to prioritize, because if you prioritize Ms. Liang's documents, you may be putting somebody else on the back burner that somebody else on your team might want. So, let's figure out among yourselves who you want to prioritize.

This request for Heideman and Soria is denied without

prejudice to reraising after we get Liang's documents.

And the same thing with the other four custodians. I don't have their names right in front of me that we talked about last time. Again, I ordered a specific production based on what was going on in California. Wait until after you've gotten those documents before you seek to expand production from those four custodians.

All right. The hit count issue.

MS. PEASLEE: Thank you, your Honor.

THE COURT: OpenAI, you seem to be almost in agreement, and I would suggest that you not get hung up on defining good cause right now, but go ahead, Ms. Ybarra.

MS. YBARRA: Thank you, your Honor. Michelle Ybarra, for OpenAI.

Yes, I think we can compromise at the 660,000 cap. You understand our position on good cause. We cannot continue to keep getting the search terms seriatim, but we take the Court's guidance, and will continue to try and work this out ourselves.

THE COURT: I don't think we're at the end of hearing about that. I expect, if I were a betting person, at least one more time, we'll talk about this, but maybe it's calming down a little bit.

Okay. Microsoft.

MR. FRAWLEY: Sorry, your Honor. Alexander Frawley,

for The New York Times. I think we have resolved the issue, for now at least, on OpenAI's search terms, but I want to make sure we're all on the same page.

So, when I heard Ms. Ybarra say that they can compromise on the roughly 600,000 document hit counts, I just want to be clear that we're referring to the January 2nd search term proposal that we made to OpenAI that yielded, I think it was, 666,000 documents.

THE COURT: Yeah, nobody wanted to say 666.

MR. FRAWLEY: Yeah, I was thinking about that.

THE COURT: Yes.

MR. FRAWLEY: With that clarification, I think that's all I have to say on this one. Thank you, your Honor.

THE COURT: Okay.

Turning to the Microsoft hit counts.

I think I'm missing a data point when I was trying to figure this out. If Microsoft saying is that they've already reviewed 600,000 documents and produced 132, is your production done?

MR. BRYANT: Jared Bryant, for Microsoft.

Absolutely not.

THE COURT: Mr. Frawley, you're at the podium. It seems like this is something that ought to be able to be worked out, but why are we in a different place with Microsoft's hit counts?

MR. FRAWLEY: Yes. This is Alex Frawley again, for The New York Times.

I agree, your Honor, that this is something I think we can work out, and we made our proposal yesterday to Microsoft to try to resolve this, and I'd like to ask your Honor to order that proposal. And I can explain it.

It's a three-step process. The first step would be that Microsoft provide us with a comprehensive list of the search terms that they have used in this case to generate the approximately 600,000 documents that they reviewed. And we have bits and pieces of that we can pull together, but we don't have a list that says, here's all the search terms we've used and the corresponding hit counts. So we think that's the most important step one — we get that list.

Then step two would be that we review that list, and we compare it against the most recent search term proposal that we've made to Microsoft. And we'll look for the gaps. We'll say, okay, here's the gaps in their search terms, they're missing these kinds of terms about these kinds of topics, or maybe they have some terms on some topics, but they're not good enough, and we'll take a hard look at those gaps. And then we'll go back to Microsoft, and we'll say, okay, maybe it doesn't have to be that you review 600,000 more documents, but it's some other number of documents based on these gaps that we'll identify for you. That's step two.

And then step three, I think — and hopefully step three doesn't have to happen — but the potential step three is that shortly from now, maybe two weeks, because, again, the theme has been setting deadlines, the parties file a joint update where we either tell the Court that we've resolved it, and Microsoft is going to review these search terms and this many more documents, or, hopefully, less likely, but possible, we submit competing proposals to the Court, where we would say here's the gaps we've identified, and we think that's going to be resolved by them reviewing this many more documents, and they say either the gaps are smaller, and we'll do something less, or they'll say none at all.

So, that's the three-step process that we proposed to Microsoft yesterday, and we'd ask your Honor to order it.

THE COURT: Okay. Well, let me hear from Mr. Bryant.

MR. BRYANT: Thank you, your Honor. Jared Bryant, for Microsoft.

Your Honor, this is the type of thing where we should have had this discussion before a motion was filed, frankly. A motion got filed, and then there's a new proposal in today that we just heard about yesterday. The Times is asking the Court to order that. I, frankly, don't think that that is necessary.

To be clear, what happened here was back in June, we provided The Times with a list of our search terms and custodians. That was on June 12th. And that kicked off the

process of negotiating proposed search terms from The Times. The Times never agreed to our search terms, and despite a lot of feedback from Microsoft on The Times' proposals that you saw in our letter, their proposals began in the millions, and we've been trying to work with them in good faith to bring those numbers down to something reasonable. The Times has never agreed to any particular search term that we ran.

So, we told them upfront what we were doing. And during this time, Microsoft didn't sit on the sidelines while the parties were negotiating The Times' search terms. We complied with our discovery obligations. We ran our own terms, we reviewed our own documents, we produced documents that are relevant and responsive in this case, and that's what's led to the 130 or so thousand that have been produced already. Clearly, we've had to review many more to get there, and, clearly, there will be many more coming. There are more documents in the queue.

So, really, what's happened is the train left the station, and there are two efforts underway. And now, after we have undertaken this effort, The Times approached us, I think just two weeks ago, with this motion asking for an arbitrary 600,000 additional documents on top of the work that we've already done.

Now, this request is not tied to any particular discovery or request for production that they claim is not

captured in our custodial search terms. It's not tied to any particular discovery request. It is an arbitrary request, long after the train has left the station, for another 600,000 documents.

Now, we agree with The Times and share the desire to bring this to a close, no question, but proposing an arbitrary cap of 600,000 documents so late in the process just doesn't get us there.

So, again, I see no reason why we can't continue to work together and get search terms resolved and produce documents that are relevant and responsive to this case. I think what The Times says here, the cap worked really well with OpenAI, so it should work well for you, and, frankly, it's just not a one size fits all solution, your Honor.

THE COURT: I agree, and I'm going to -- in the interests of time, because this is still getting longer, are you saying that you've already provided the news plaintiffs the list of search terms that you ran to generate the 600,000?

MR. BRYANT: At the outset, we provided them, as we're required to do under the ESI order, our search terms.

Now, throughout discovery, we've continually added custodians and run those search terms against custodians. What has not happened, and I don't think it's, frankly, required, is providing The Times with realtime updates, like, each and every time a custodian is added — here's what's in our queue. But

the bottom line is it's no secret that to produce 130,000 some documents, we're reviewing a whole lot more.

THE COURT: Okay. So, what's already been reviewed and produced that The Times has, that's that June list of search terms, you ran those?

MR. BRYANT: That was the beginning. There's more than just the June list. That was the genesis and the basis.

THE COURT: Okay. Can you put together quickly all the search terms? So I'm not going to order it, but I'm going to ask you to continue the meet-and-confer process, confirm the list of the search terms, so that The New York Times can put that up against their search term proposal. I'm not going to order a number of hit counts, but I expect you to work in good faith to see, as Mr. Frawley said, are there -- identify if there are gaps. Maybe there aren't gaps, and that's okay, too. We can move on to depositions.

MR. BRYANT: And, frankly, that's all we want, your Honor. We just want something tied to the discovery requests in the case.

THE COURT: Yes, yes.

So, have that meet-and-confer in the next two weeks. Get the list of terms. You may believe you've provided them already. If you've --

MR. BRYANT: We can identify where it is, and we have --

THE COURT: Just put it altogether in one, send it to The Times, have the meet-and-confers in the next two weeks about whether there are gaps, and think about what you might want to do about those gaps.

I'm not going to order new hit counts — it sounds like you've done a lot of the review — but I think what I was starting with this was that I didn't have an understanding of what sort of led to this dispute showing up this way. Now I have some understanding about it.

This is the filling in the gaps exercise. This is not like make new search terms and reinvent the wheel here, okay?

MR. BRYANT: That sounds fine, your Honor. We appreciate that.

THE COURT: Okay.

MR. BRYANT: Thank you.

THE COURT: All right. Thank you.

Okay. Next on my list was the four custodians, which I just talked about, and then after that is ECF 398 and 419, which seems to be related to damages discovery and also some profit and market analyses issues. It seems very broad.

So, Mr. Frawley.

MR. FRAWLEY: Alexander Frawley, for The New York Times.

And, yes, Docket 398 is a motion that we filed against OpenAI about document disputes. And there's really three

categories of document disputes.

The first category, I'm happy to report that we resolved. That first category was documents concerning the market for training data and retrieval-augmented generation data. So we reached an agreement where OpenAI is going to do a search for those documents. So, that issue is resolved.

THE COURT: Yeah.

MR. FRAWLEY: The two other issues are not resolved, and I will start with the next issue, which is documents regarding OpenAI's revenues and valuation.

And to take a step back, we're talking about a very young company here in OpenAI, but one that recently secured over $6 billion in investments, and they're valued at 157 billion. OpenAI has said to us, well, all you really need are our financial statements. But we don't think financial statements, or they might not paint the full picture of where OpenAI is financially and where they're going.

And, indeed, what OpenAI is pitching for the kinds of investments that they're securing, we very much doubt that they're relying exclusively on their financial statements.

So we're seeking documents and presentations about OpenAI's current and projected revenues and their valuation. And we think those are relevant to at least the fair use analysis, as well as damages. I'll start with fair use.

For factor 1, for the purpose and character of the

use, courts have been directed to consider "whether the allegedly infringing use is commercial or noncommercial." And that's the *Napster* Ninth Circuit case, 239 F.3d at 1015. And unless OpenAI is willing to stipulate that they've entirely abandoned their original nonprofit vision, and admit that they're now, going forward, going to focus exclusively on commercializing their products and making money, we think we are entitled to discovery to understand not only how they're making money today, but what their plans are going forward, and, in particular, how they're going to use the products that they built with our content, improperly, to continue to become a commercial company.

We think the information is also relevant to damages. The alleged infringement in this case is ongoing. So we're not only entitled to what OpenAI has already made from its infringement, but what it continues to make every day, and what it's going to make. And we cited the *Looney* case, 2010 WL 5175167 at *2 for this, that future profits are recoverable. And we understand OpenAI's objection here to be, well, future profits, future revenues, those might be speculative, but that's no reason to refuse this discovery. That's an issue for down the road. Down the road, OpenAI might argue, whether at summary judgment or in a *Daubert* motion, that we've put together some requests for ongoing or future revenues that are too speculative.

But that's, again, for down the road. All we're seeking now is the discovery, which is narrowly tailored to a discrete set of documents. And I think OpenAI cited a case in their brief, which actually makes our point pretty clear and supports our request for this discovery. And that's the *Burns v. Imagine Films* case, 2001 WL 34059379. And this was in OpenAI's brief. And that court was also considering a request in a copyright case for future profits. What the court said was, "Certainly, the term 'any profits' in Section 504(b) can be read as encompassing future profits." And the court went on, "that the revenues have not yet been received is not in itself a bar to arriving at a reasonably accurate calculation."

Now, in that case, this was a summary judgment decision, and the court refused or denied the request for future profits as overly speculative, but the reason was interesting. That was a case where the defendant had made a movie allegedly where parts of it were infringing on the defendant's content, and the defendant said, well, if we lose this case, and if we were found to have infringed, we'll just take out the parts of our movie that were infringing, and we'll remake the movie without those infringing parts. And that's the basis on which the court said, okay, fine, the future profits are going to be speculative.

So, that case is different for two main reasons. One, again, we're not at the summary judgment stage, so this concern

about speculative profits is premature; and point two, OpenAI has never said that if they're found to have infringed, that they're going to extract The Times and Daily News and CIR content from their models, from their products, and then basically redo them without our content. So that's another reason why, even down the road — again, it's too early to really be doing this analysis — but down the road, we'd be able to distinguish that case.

So, we think these requests are relevant, and we'd ask your Honor to order them.

That's one category of documents. It might make sense to stop there if you have questions.

THE COURT: Actually, get the third category out.

MR. FRAWLEY: Sure.

MR. SUN: One clarification. Christopher Sun, on behalf of OpenAI.

I'll be handling the argument on the question of the financials, but a different lawyer is going to be handling the argument on the last category. I just raise that in case it might be more helpful to segment them, but I defer to your Honor.

THE COURT: I just want to get it out.

MR. SUN: Sure.

MR. FRAWLEY: The next and remaining category in this is documents produced to governmental authorities, and we're

asking OpenAI to identify any foreign governmental authorities to whom they have produced documents or given information that's relevant to the factual allegations in this case.

What the dispute is here is that OpenAI has refused to even investigate this question and tell us whether there's any foreign governmental authorities to whom they've produced information that's going to be relevant to this case, and their argument is, well, if it's a foreign governmental authority, they might be applying non-U.S. law, and how could that be relevant. But what we're seeking are the documents that would be given to these foreign governmental entities, and those documents are going to be factual in nature, or responses to questions --

THE COURT: Yes, so I'm going to cut you off here because this is where I got lost. But why? Why are documents produced to foreign governmental entities relevant for this case? And part of that relevance question is also, and why would they not be duplicative of what's being produced here anyway?

MR. FRAWLEY: So, they may be duplicative of what's being produced here, because I agree that, to the extent there's going to be documents produced that are relevant to issues in this case, we've probably served requests for those kinds of documents. I think a good example is there was a dispute earlier on in this case, in the Authors Guild case, to

the FTC interrogatories, and that's information that was provided in an interrogatory form to the FTC, nonpublic interrogatory responses, that eventually OpenAI agreed to produce in this case.

So, those were documents that weren't being given otherwise with respect to the requests that had been served. And we don't know if there's a similar kind of document for a foreign governmental authority because they haven't told us whether there is or isn't, and there may not be, in which case this is a pretty easy dispute to resolve.

THE COURT: But why foreign governments? I can see for FTC, we're dealing with the United States, right, and this case in the U.S. involving U.S. law, but why foreign governmental authorities?

MR. FRAWLEY: So, we described one example, your Honor, in our brief, which was a submission that OpenAI made to the House of Lords in the U.K., and that was a submission where they provided information about how they train their models. So, however that was being evaluated under maybe a different set of laws, whether that's copyright infringement or not, it's still the factual information about how those models were being trained and how OpenAI was or wasn't using particular content. So those kinds of documents or those kinds of information could be relevant here. We just don't know if that kind of information has been provided because they haven't told us.

THE COURT: But you ought to be getting it in the discovery here, right? You're getting the information. You're not getting the submission to the House of Lords, but you're getting the information, right, the underlying information you're seeking because it's relevant to the case.

MR. FRAWLEY: And that one, your Honor, is public, so we do have that one. That's how we know about it. But we might be -- I agree that we should be getting sort of the underlying technical documents already with our request. It's really -- you know, the FTC, I think, is a good example because it was a submission that otherwise wasn't being provided, and we just don't if there's something like that for any foreign governmental authorities.

THE COURT: Yes, I get that. But I'm not convinced that you get it because it's going to a foreign governmental authority.

MR. FRAWLEY: And maybe, your Honor, what the issue here is, and the first step could just be, they tell us whether there are any kinds of situations like this, which we don't know, and if there are, and they say here's what there is, but you're not getting all of it, we don't think this is all relevant, it would still help move the discussion forward to where we get -- you know, maybe request a smaller piece of it. We just think a first step would be that they actually tell us whether these sorts of proceedings are occurring where they are

producing this kind of information.

THE COURT: I'm still not seeing it. I'm not seeing why -- I'm not seeing why you need it, but let me hear from OpenAI on this third issue first.

MR. DAWSON: Thank you, your Honor. Andrew Dawson, for OpenAI.

I won't say much because I agree with your Honor's observation. The one observation I would add is you commented a moment ago that this is the time not to reinvent the wheel, but to fill in gaps. Discovery has been pending for months. The Times has issued 130 RFPs. The parties have hammered out and meet-and-conferred the scope of responsive documents. Those issues have been litigated here. Now to reopen the door widely to an entirely different set and to outsource discovery to government regulators abroad, there's no justification for it. All relevant information, The Times has been able to craft their discovery requests directly, OpenAI is making those productions, and there's no justification to go broader.

And I'd be happy to answer any questions you might have.

THE COURT: Well, I think I'm going to deny this one. Again, I'm not telling you how to litigate the case, but I also see a lot of 1782 petitions here for discovery in aid of foreign proceedings, but I'm not seeing it as something that's core and necessary. If you're trying to get the information

and statements that OpenAI has made to others and other governmental entities, I think there are other ways that you can get that. But I'm not going to go there now.

MR. DAWSON: Thank you, your Honor.

THE COURT: All right. More concerning is the revenues and valuation.

MR. SUN: Christopher Sun, on behalf of OpenAI.

Your Honor, I think there are two components to this financial request, and it might help to clarify the two different categories.

The first category would be, broadly speaking, I think they're requesting presentations that discuss discrete categories of financial information, like revenues and valuations. And I recall that your initial comment with respect to this dispute was that the request seemed broad, and I thought it was telling that Mr. Frawley did not dispute that characterization.

Instead, he kind of made a bunch of arguments that I think prove the point. He said that the financial statements might not paint the full picture, and that there might be missing information, and then he referenced OpenAI potentially pitching investments. But what was missing from Mr. Frawley's explanation, and what has been missing, frankly, throughout the entire meet-and-confer process, is some identification of what information he thinks is missing or what relevant issue the

documents that OpenAI has produced or otherwise agreed to produce don't already address, because I think they do. I think we've already offered to produce all of the information that they ultimately need.

And, of course, if there is information that we haven't agreed to produce yet, we'd be willing to meet and confer about it, and we can meet and confer about the best way to identify and produce that specific category of relevant and responsive information, but that's not what they're asking for here, your Honor. They're asking for, essentially, any presentation that contains the word revenue, and it's unclear to me why it would be appropriate to conduct such a broad search.

An example that comes straight from their brief is they want -- I think they would be asking for any presentation that contains the revenue figure from one of our years, and that figure is obviously in the audited financial statements that we've agreed to produce.

THE COURT: Okay. So you haven't produced the audited financial statements yet.

MR. SUN: We have, your Honor. That's why this is befuddling, because we have produced the audited financial statements. If there is a question that the pages and pages of notes that that financial statement does not answer, we are happy to meet and confer with them about what else they need.

THE COURT: Well, talk to me about why -- the fair use is a defense, right, it's your defense.

MR. SUN: Oh, yes, your Honor. I think the --

THE COURT: The first fair use factor.

MR. SUN: So, the first fair use factor is infringing use, and I think the operative word with respect to this is the nature of the infringing use. What Mr. Frawley referenced exclusively in his argument was about the commercial or noncommercial use by OpenAI, the company as a whole, and the case law is very clear on this, the *Phoenix Tech* case that we cite in our brief references this, but the relevant question in terms of factor one is the use, whether or not the use is commercial. And I reiterate that we've produced financial statements that speak to the revenue generated by OpenAI's variety of products. We've offered to produce segmented revenue information for our various products and models, to the extent that information is kept in the ordinary course of business. So it's entirely unclear to me what additional information could be provided by these unspecified presentations that wouldn't address that commercial use factor.

THE COURT: Well, I guess my question, and I am going to ask Mr. Frawley to clarify, because, as I understand, the audited financial statements are looking back, right? This is back to the, like, looking back versus looking forward, right? So, looking back each year is whether the use has been

commercial?

MR. SUN: There's two points. I think we're pivoting to the second point, so I'll just signpost that we're there, which is I think the second category of documents they're requesting are the projections.

With respect to factor one, I'm not sure why -- I think there are other categories of documents we've offered to produce on this question, but the question of whether or not there will be infringing use in the future, I don't believe, is a relevant one for liability in terms of fair use. The question is whether or not the use in the past was fair, and we've produced documents about that.

Also --

THE COURT: No, I think we're talking past each other.

MR. SUN: Sure.

THE COURT: I think the concern — and I'm going to ask Mr. Frawley to clarify — is, if it's commercial — and he made this remark, you know, that OpenAI is still a relatively young company — that commercialization isn't just what you make off of the product, but potentially what fundraising might be coming in. And that's why I see potentially projections, as well as presentations, and all of that as potentially being relevant. I'm not giving it to them. I want you to explain to me what, and then I want Mr. Frawley to explain to me why, the categories of the documents you're seeking, why the plaintiffs

are seeking now, why the financial statements aren't enough.

MR. SUN: I guess, your Honor, I would have two responses to that. I'm aware of no law, and plaintiffs have cited no law, that suggests that investments would be relevant to the commercial use inquiry.

And the second argument that I'd want to make is, again, that seems like we're getting very far afield to the question of the nature of the use, whether the nature of the use was fair. If we extend the number of that to include things like whether or not a company can get investments, whether that speaks to the commercial nature, we're expanding the inquiry far beyond kind of my understanding of what that inquiry otherwise encompasses, but also the kind of tailored approach to use. They're focused on use.

The *A&M v. Napster* case that Mr. Frawley references specifically says that the inquiry should be focused on use. For example, it says the purpose and character element also requires a district court to determine whether the allegedly infringing use is commercial versus noncommercial, not whether or not the company as a whole is operating in a commercial capacity, not whether or not the company as a whole is to garner any investments, because that would really blow up the question of fair use.

But the other thing I wanted to highlight before we got lost, and before I forgot, is that Mr. Frawley

strategically focused his discussion of the case law on two particular cases, which were *Looney* and *Burns*, and I am happy to discuss why he's wrong about *Burns*, but I figured it would be worth pointing out that we have cited only two cases on the case of whether or not projections are relevant in a copyright action. Both cases come out of the Northern District. One came out this month from Judge --

THE COURT: The Northern District of?

MR. SUN: California.

Both of those cases found, unequivocally, that projections are not relevant and not discoverable in a copyright case specifically with respect to damages. And that's law that the plaintiffs have never been able to dispute. They've found no contrary cases. And, in fact, the *Looney Ricks* case that they cited proves the point that this is probably the state of the law because the judge in that case did a survey of, like, laws around the country, and he was -- or she was specifically inquiring about whether or not the statute would allow for the collection of future profits, but they were unable to identify any kind of law on the question. So I think OpenAI has cited and canvassed the law on this question, and it is unequivocal that projections are not relevant in a copyright case, but specifically to damages.

And to address the *Burns* argument, your Honor, I think there are two clarifications to Mr. Frawley's point.

The first is that *Burns* wasn't relying specifically on the question of just whether or not there would be a change in the movie for purposes going forward if infringement was to be found, but, rather, the crux of the court's opinion there was that plaintiffs had received the actual sales and the actual financials, the historical financials, so there was no need for projections, which would likely be speculative. So that's the main thrust of the court's reasoning.

But with respect to changing the nature of the film, I think that reason applies equally here. Mr. Frawley and the plaintiffs are asking for projections going many years into the future, and for those to be relevant, the assumption has to be that if there was a judgment in this case, if this case goes to trial and there was a judgment, they are using projections that would have been prepared before that judgment issued. And, thus, if a judgment did issue, they'd be basing any damages calculation on the assumption that following a presumed -- presuming they think they're going to win, and they're calculating damages based off the assumption they're going to win, that they are using projections that don't take into account that judgment and, therefore, assuming that OpenAI would have received a judgment of infringement and then done nothing to alter any of its practices or its products going forward. That's why it's speculative because it's based in assumptions that will not be true when the damages model is

ultimately collected. And that reasoning, with regards to what Mr. Frawley characterized as the *Burns* opinion as addressing, that reasoning applies here as well.

THE COURT: Okay.

Let's hear from Mr. Frawley.

MR. FRAWLEY: Your Honor, I'll just briefly start with case law.

I'm not sure how OpenAI can say that the case law unequivocally supports them when the case they cited supports us from the quote that I read. I don't think I need to read it again. So I'll leave the case law at that.

In terms of your question, we got the financial statements, why aren't those good enough, and I think your Honor hit the nail on the head when you were talking this through, which was the financial statements are capturing a moment in time and costs, revenues. And, remember, this is a very young company, and often for startup type companies, at the beginning of their tenure, they don't initially turn a profit, and their evaluation and their plans are based on what they expect to do in the future, their projections.

So, financial statements won't cover projections. It won't be the reason that others want to invest in the company, right? The people want to invest not because of what they're doing right now, but because of what they think you can do going forward. And that's the kind of information we're

seeking, those sorts of revenue projections, and we think that's going to be -- and I'll just make, I think, one more point, your Honor, which is that this is, in some ways, a unique case because the main product that's at issue, from OpenAI's perspective, is ChatGPT, the user facing chatbot they released barely over two years ago. And in this case, at least at this point, most of their commercialization, most of their revenues, are going to be all revolving around that particular product and plans for that particular product. And that's a product that we very much allege has been improperly built with news plaintiffs' content.

So, we're seeking to understand the profits and the revenues that are projected from that. That's relevant to damages. But also for fair use. OpenAI talked a lot about how we have to focus on the use, the use. But, here, what does the use mean? The use means, what are their plans for ChatGPT, the main product at issue in the case as to OpenAI? How are they going to continue to commercialize ChatGPT? What exactly will be the plans for ChatGPT? What will it focus on? What sorts of -- what's going to be the specific areas that they want to improve upon? Is it going to be news content, is it going to be something else? And those are the plans that we think are relevant to the fair use analysis and, in particular, the commercial use prong, the first factor of the fair use analysis.

So, your Honor -- also, I have one final point, because I think counsel for OpenAI suggested the request is overbroad, and that we are basically asking them to produce any single presentation in OpenAI's files that hit on any kind of revenue figure. And that's not what we're asking. In terms of custodial documents, these kinds of documents are going to come up in the search terms that we've already proposed, and that, as of today, are now the finalized subject of good cause, search terms for their custodial searches. So, we just ask them to produce any responsive documents that come up through that custodial search process.

For the noncustodial part of it, no, we're not asking them to go track down every single presentation. We're asking them to undertake a reasonable investigation for projections and presentations focused on OpenAI's revenues. And we'd be surprised to hear if that's going to be a very burdensome undertaking. We don't think it would be.

THE COURT: So, let me go back to what you just said, which is you were asking for those documents that turn up in the custodial searches. Has OpenAI said they're not producing them?

MR. FRAWLEY: That's my understanding, yes. And that's only a piece of it, your Honor, but that resolves one piece of it, now that we have -- we now have a set of search terms as of a few moments ago.

But we'd also -- I just want to make clear that we think it's not limited to custodial documents that will come up in the custodial search. We think it should also be a reasonable investigation for noncustodial documents. And, in particular, we've talked about projections, we've talked about investments. We don't think these kinds of documents are going to be ones that are going to be particularly hard to find if they undertake a reasonable investigation.

MR. SUN: Your Honor, can I respond?

THE COURT: No, not yet.

MR. SUN: All right.

THE COURT: I tend to think that this is a little bit premature, but I will take a look at the briefing and at this transcript. I'm not going to order that OpenAI produce or not produce certain categories, but, in the meantime, I really encourage you to think about whether there are particular investments or transactions or particular presentations or other documents that you might get, or that you might be getting, or that you might want to ask OpenAI to focus on. In other words, try to address OpenAI's concern about the request being overbroad by not asking for all of this, but ask for a subset of it, see what you get, see if it leads anywhere.

Also, see if any of this comes out in the production that you are getting. If there is -- I'm just making this up, but if there are documents that you do get that hit on the

search terms that start talking about plans for the future, that seems to me that that's all part and parcel of the production and the discovery that you're going to be getting anyway.

MR. FRAWLEY: Can I ask one clarification question?

THE COURT: Yes.

MR. FRAWLEY: So, it sounds like, based on what your Honor just said, that for custodial documents that are going to be coming through the custodial review, that you think OpenAI should produce documents that are responsive to these RFPs at issue in this motion and the issues that we've been addressing today.

Is that correct? So, in other words, these are documents that they're not separately having to go find because they're part of the custodial searches, and the only direction is that, if you come upon, through the searches, a document that is responsive to these requests, that it should be produced?

THE COURT: I'm already -- I feel like I'm already too involved in your document review of production and how you plan for your depositions. I get my hands too much in here.

I understand that you've got search terms that you've agreed upon, and you're going to work on narrowing those search terms to get to that 666,000-ish hits.

MR. FRAWLEY: We're already there.

THE COURT: Right, right.

So, in those documents, are there going to be hits that get these documents, these particular documents, that you're looking for?

MR. FRAWLEY: Yes, there will be. The only problem is if OpenAI says, oh, this document is about financial projections, but we don't think that's relevant, so we're --

THE COURT: I see Mr. Sun shaking his head. So maybe now would be a time for you to speak.

MR. SUN: I'm not candidly sure what Mr. Frawley is talking about. The first time I heard about this proposal regarding, here's what the custodial search would look like, here's what the noncustodial search would look like, is right now in his responsive arguments. This is the kind of thing that I think it would be, to your point, more productive to discuss in a meet-and-confer, and we're happy to do that. I think that it sounds like Mr. Frawley is willing to make modifications that will move this negotiation forward.

I don't think the right thing to do is to have a meet-and-confer in open court, when I haven't had an opportunity to kind of check with the team doing these productions, doing the review, checking with the client, and asking for advisory opinions about, well, what if we offered this, that would be reasonable, wouldn't it? That's something that we can all work out together. I'm happy to do that with

Mr. Frawley.

THE COURT: Okay. I started this whole conversation thinking that this dispute might be a little bit premature because I want you to get through that document production, I want you to have your meet-and-confers about, hey, we expected to see this or this or this, are you holding this back, well, what about, you know, this reference in the documents. That's all a meet-and-confer. We shouldn't be having it in court, we shouldn't be having it the day before, you know, you have to submit your charts, but -- and I'm not going to tell you how to -- I'm really, really trying not to micromanage your discovery, even though it sounds like I am.

So, please go back and see what the review turns up, right, of the custodial review. What does it turn up? Are there documents that are responsive to these requests? I don't think that they should be withheld in bad faith. You know what the issue is. You get a lot of extra stuff, maybe it's not a hundred percent relevant, but you don't have the time for this, right, you don't have the time to pull out documents and be like, oh, this is not good for us, we're going to hide this. And I don't expect anybody would be doing that. And I'm not sure there's a dispute here.

MR. SUN: For what it's worth, your Honor, I agree. And I agree that we could resolve this, and should resolve this, in a meet-and-confer.

MR. FRAWLEY: And, sorry, really quickly, your Honor?

THE COURT: Okay.

MR. FRAWLEY: And, your Honor, we're happy to meet and confer.

To guide the meet-and-confer, as long as we can agree that financial projections of the kind we're seeking are relevant, that would be --

THE COURT: Go talk. Go talk not here, not on the record and not on the transcript.

MR. FRAWLEY: Thank you, your Honor.

THE COURT: All right.

ECF 322, I think, is not in dispute now.

This issue with RFP 2, was that resolved by our conversation just now?

MR. FRAWLEY: Alex Frawley again, for The Times.

It's a similar issue to the RFP 106 that we talked about. RFP 2 is kind of the analogous RFP, but focusing on U.S. entities, not foreign entities.

The issue here is that OpenAI has said there are no responsive documents to the RFP aside from the FTC responses they've already produced.

THE COURT: So why isn't that the end of it?

MR. FRAWLEY: It could be. What we brief in the chart, your Honor, is that we have raised a few examples to them of investigations that are publicly reported. And one is

a Delaware investigation into -- or a potential investigation into their for-profit transition, and we've said what about this one, have you produced any information as part of this proceeding that could be relevant to this case? And their response to that was, no, that proceeding is not relevant; we don't have to talk about that proceeding.

And that's the tricky part, because, as you know, the parties have many times had disputes about what is and isn't relevant. So, when we're seeking relevant documents produced to a governmental entity, there is going to be gateway disputes about, well, was that investigation relevant or not. So that's what we briefed in the chart, your Honor. I don't think I need to belabor the point, but that's where we are on RFP 2.

THE COURT: All right. This one does seem to shade into some of what we were discussing earlier. I'm not -- I suggest you continue to meet and confer. We've heard a little bit about fair use and commercialization today, and I don't have any visibility into this, but you all should be talking about that. If this bubbles up into another motion, it will, but I want you to get a little bit clearer on that and be able to tell me why, for example, you need documents that you're not already getting or documents, proximate documents, that you are not already getting from the document review.

The other potential concern about agencies and governmental investigations and things is that there could be

other restrictions on why you get documents, and maybe the information you get is through a different discovery process.

Mr. Dawson, you are standing up. You want to say something?

MR. DAWSON: I don't have much to say. I think the meet-and-confer process might be useful. The one thing I'll say is, direct RFPs to OpenAI, we can hammer out relevance, we can discuss, we can bring those disputes to your Honor. When we're going the circuitous third-party route, that is disruptive. That is, we can't talk directly, we can't negotiate directly, and there's a third party with different interests, different rules. It's better to just proceed directly. We'll obviously discuss -- I know there are direct RFPs on these subjects. I'm sure we'll continue to meet and confer on those, but I'm hesitant, given that circuitous third-party aspect.

THE COURT: What do you mean by "circuitous third-party aspect"?

MR. DAWSON: Because if a document is produced to a regulator, OpenAI and The Times aren't in the room to discuss what's relevant and what's not. We are in the room here, we're in the room in meet-and-confers, and a third-party that has its own processes, its own incentives, its own statutes, its own everything, it just adds very old stuff that we don't need to address if we're face to face.

THE COURT: Yes. And they may have a different focus. So you could be getting some things that you are interested in, but maybe not all of the things that might be useful for you.

Continue to meet and confer.

Document-related disputes. Microsoft, that's ECF 396.

We're almost done, I promise.

MR. FRAWLEY: It's me again, your Honor.

THE COURT: Yes.

MR. FRAWLEY: Alex Frawley, for The New York Times.

So, there are three categories of documents that are in dispute. And I want to start with the third category because it's pretty quick.

This third category is documents concerning the impact of Microsoft's conduct in this case on content creators. And the problem here that we briefed was that for noncustodial documents, Microsoft was previously limiting its investigation to documents that specifically addressed the impact on the three specific news plaintiffs.

And we understand now, from Microsoft's opposition brief, that they're willing to do a broader scope, and that they're not going to limit the investigation to the three news plaintiffs. So, I think the parties should meet and confer on what that exact scope is, as long as I'm correct that Microsoft is no longer standing on the limitation to the three news plaintiffs.

THE COURT: Well, meet and confer about that.

All right. Number two?

MR. FRAWLEY: Number two, your Honor, is probably the meatiest dispute here, and this is the dispute about documents concerning the market for training data and retrieval-augmented generation data. And the crux of this dispute is that we're seeking documents and communications concerning Microsoft's negotiations over license agreements, and Microsoft's refused to provide -- I'm sorry, I should be clear. Licensing negotiations that did not result in a final agreement. And Microsoft's objection to producing these documents is primarily that they think your Honor foreclosed this route of discovery in a ruling in the Authors Guild case that sought licensing documents, but we think that order is distinguishable, and I'd like to explain why.

THE COURT: Okay.

MR. FRAWLEY: With that order — and we saw this in the briefing in Authors Guild, we saw it in the argument in this court in December, and then we also saw it with your Honor's ruling — that was driven by Microsoft's concern that what the class plaintiffs were seeking, in Microsoft's view, was discovery into models and products that were not at issue in the case. And they said in their brief, and I will quote, "Because Microsoft's LLMs are irrelevant, so are documents related to unexecuted data access agreements." And that's

Docket Authors Guild 279 at 1.

And then Microsoft made the same argument before this Court at the hearing. They said the plaintiffs were seeking -- sorry -- the plaintiffs were seeking documents about "negotiations regarding data access for Microsoft's LLMs," and they went on to say that implicates "an entirety different set of LLMs that are not at issue." That's the December 3rd transcript, at page 138, line 2, to page 139, line 7.

And then this Court's ruling was tied to that issue. This Court said that you would deny the motion "for the same reasons as articulated in ruling on ECF No. 279 in the newspaper cases." That's the same transcript, at 142, 7 to 9.

And ECF 279 in the News case, in our case, was our request for documents concerning Microsoft's anticipated products. So your Honor said, at this time, we're not going to expand the scope of discovery, the news plaintiffs aren't getting anticipated products from Microsoft at this time, and the Authors Guild plaintiffs aren't getting data accessing agreements or negotiations about Microsoft products that aren't yet subject to discovery in the case.

That makes our request now, all that background, a very different request, and a much narrower request, because what we're seeking are licensing negotiations for the practices and the products that have always been indisputably relevant to discovery in this case. And, specifically, I'm talking about

retrieval-augmented generation — RAG for short — and the main product here, from Microsoft's perspective, is Microsoft Copilot, its user facing chatbot. So we're seeking licensing negotiations for how Microsoft uses content or has sought to license content for RAG and for use with its Copilot chatbot.

And we think those communications are relevant. We don't think they're covered by the Authors Guild ruling. And here's how we know that they're not covered by the Authors Guild ruling — because Microsoft has repeatedly distinguished the News cases from the Authors cases by arguing that in the News cases, RAG and Copilot are relevant, and in Microsoft's view, those things aren't relevant in the Authors Guild cases.

So, we don't think your Honor's ruling could have applied to RAG and could have applied to Copilot, and our request here, given that Microsoft has admitted in our case those are relevant practices, those are relevant products, and that's the focus of our request for these licensing negotiations.

I'll stop there if your Honor has any questions about why we think the Authors Guild ruling is distinguishable.

THE COURT: Okay. Let me hear from Microsoft.

MR. BRYANT: Thank you, your Honor. Jared Brian, for Microsoft.

On this issue regarding license negotiations, there are two arguments that we're making: Number one is that this

issue was raised and decided. Number two, even if it weren't, what The Times is seeking above and beyond what we've already agreed to produce is not relevant. So, let me take both of those in turn.

The request for production The Times is seeking to compel in this letter motion is RFP 61. It's set forth in Footnote 3 of their brief. And RFP 61 seeks documents and communications concerning negotiations between, I'll say, Microsoft and owners of copyrighted content, including The Times, over Microsoft's use of copyrighted content in, I'm going to say, their products and services.

This is nearly a verbatim request to the RFP that the class plaintiffs were seeking to compel. The request for production the class plaintiffs sought to compel at the last hearing was class plaintiffs' Request For Production 46. This was referenced in the class plaintiffs' motion at ECF 269. That's RFP documents sufficient to show attempts to obtain a license for use of content in connection with any LLM, including negotiations.

So, the issue that the class plaintiffs were seeking to compel were negotiations regardless of whether there was a meeting of the minds, regardless of whether they resulted in an executed agreement, with respect to any LLM. So, that is nearly verbatim the same request that the news plaintiffs seek to compel here in RFP 61. But here's the thing: Even if it

weren't already decided on this issue, what Microsoft has already agreed to produce — and we mentioned this at the last status conference — is, number one, any executed agreements; number two, the negotiations leading to those executed agreements, because we agree they might have some relevance to the question of the existence of market; and then, in addition to that, we have also agreed to produce any of the negotiations with The New York Times, which obviously did not lead to an executed agreement, or negotiations with any of the other news plaintiffs.

Anything else beyond that -- so, clearly, we're not fighting about that. And so, instead, what we're fighting about, and what the dispute is, is about any negotiation with any content creator, anything that might come into the door, to Microsoft, regardless of whether it led to an agreement. And that's the second argument we're making, your Honor, is that these discussions, these negotiations, don't lead to any sort of meeting of the minds, they don't lead to any sort of setting of value for purposes of determining the existence of the market. So, they're not relevant in this case. And we made that argument in the class motion, and we're making that argument here in our brief.

So, it's really twofold. Number one has already been decided.

Number two, even if it weren't, the documents beyond

what we've already agreed to produce, namely, executed agreements, negotiations leading to executed agreements, and then negotiations with any one of the news plaintiffs, it's just not relevant in this case because it doesn't lead to make the existence of a market more or less likely because, frankly, there's no meeting of the minds.

MR. FRAWLEY: Briefly, your Honor, can I respond to that?

THE COURT: Yes, please do.

I really do want to focus on the unexecuted agreements or the negotiations, because my concern is you're asking to produce a negative, and that's why I'm not -- and I'm trying to understand how that's relevant to a claim or a defense in this case.

MR. FRAWLEY: Sure. And that's what I wanted to address, your Honor.

So, we're seeking the kinds of communications -- maybe I can start with an example. If there is a negotiation between Microsoft and another news provider where Microsoft was sending whether it's emails or draft term sheets, and they said, we're willing to pay you X dollars to use your news content in Copilot, even if there was never a deal reached, maybe because the news content provider wanted more money or just decided that they didn't want to be a part of this, Microsoft's admission that they were willing to pay money to use the

content for the practices and products at issue in this case for a news organization, that would be relevant to the existence of the market to explain what the market looks like, to what our potential license would look like. So we think those would be squarely relevant.

And if I can point your Honor — it's cited in our brief — to the *Hatchette* case, discussing fair use factor four. This is the Second Circuit case, 115 F.4th 163. And the Court said, "It is indisputable that, as a general matter, a copyright holder is entitled to demand a royalty for licensing others to use its copyrighted work, and that the impact on potential licensing revenues is a proper subject for consideration in assessing the fourth factor."

And the Court went on, "When we are looking at that impact on potential licensing revenues, we look at traditional markets, reasonable markets, or likely to be developed markets." So, when there's negotiations in this relatively new market around how we're going to use content in Copilot, how are we going to use content for RAG, that's relevant to establishing the existence or the likely to be developed market, and it's relevant for showing what that market looks like, what the prices are, especially when we're seeing, or we would expect to see, offers from Microsoft.

So, we think those communications are going to be squarely relevant even where Microsoft and its counterparty did

not ultimately reach an agreement.

Now, I heard Mr. Bryant had a concern, which I think is a reasonable one, which is, what happens when they just get a random email soliciting, hey, do you want to buy our content, like where does this end? And I agree that there has to be some kind of scope. And we actually worked through a similar issue with OpenAI. We filed a motion way back when, it was RFP 6 to OpenAI, about licensing negotiations, and we ended up meeting and conferring and resolving this.

I'm not saying it has to be the same way we resolve it, but there's a way to meet and confer about this to figure out the scope. Okay, it's not executed licenses, it's something short of that, but it's also not, we got one random email from a potential counterparty. So we think there's a way we can meet and confer about the scope.

But we do think that these negotiations about unexecuted agreements are going to be relevant, and we ask your Honor to at least agree with us on that, and then we can hash out the scope together.

THE COURT: I'm going to tell you to meet and confer. Here's what I am going to tell you to meet and confer about: Whether there is a discovery mechanism where you might be able to find some information about whether there's other news organizations that aren't the plaintiffs in this case, and that may be some lifting on your part, also, on the news plaintiffs'

part.

And I will leave it at that — meet and confer. If there is going to be some production or some exchange of information, I expect it to be more narrowed and more reeled in, in that it gives Microsoft some understanding of what they're looking for.

And if you're talking about a developing market, I think that you should narrow the scope of that meet-and-confer to what the market might be in the News cases.

MR. FRAWLEY: Understood, your Honor.

THE COURT: Okay. Or at least start with that.

MR. FRAWLEY: And, your Honor, if I may, that was a big part of this motion at Docket 396, but there are a few other parts.

THE COURT: Okay.

MR. FRAWLEY: I think that was the hard part.

THE COURT: Okay. Thank you.

MR. FRAWLEY: So there's one other sort of, I think, very quick issue, and I think it's quick because I think the parties are reaching agreement, or close to agreement.

We also, in this motion, said that we'd ask Microsoft to search for and produce documents that address the market for this data more generally, even if the document is not referencing a particular negotiation with a particular counterparty. And from Microsoft's opposition brief, it seems

like we're getting close here, and that Microsoft isn't objecting to considering search terms and running search terms to target those kinds of documents.

So, if I'm correct about that, then I think this issue, this subissue, can be put to bed, as long as I'm correct about that.

(Continued on next page)

THE COURT: This is part of the whole, like, get the search terms and fill in the gaps.

MR. FRAWLEY: It can be, your Honor. Exactly.

THE COURT: Yes, it is. It is. We're done.

OK. We're done with that part.

MR. FRAWLEY: We're done with that part.

There is one more category at issue in this docket 396 motion, and that's documents concerning Microsoft's investments in OpenAI and the disputes about RFP 123, which seeks documenting concerning whether Microsoft's potential or actual investments in OpenAI, including in the years 2016, 2019, and 2024, would affect OpenAI's nonprofit status, mission, or ability to engage in noncommercial activities.

We talked a lot about the commercial stuff today. I don't think I need to say anything more about that. But here is what I to want say about this dispute with Microsoft. So Microsoft's response -- and they said this in their RFP response and they said this in their brief -- they said we already agreed to produce documents about out investments in OpenAI in response to prior RFPs.

So what is missing? To the extent that anything on this RFP, you know, exists is going to be in what we already agreed to produce, at least that's my understanding of the argument. And our response to that is this has kind of enabled to bubble up a very concrete dispute about the timeline of

Microsoft's investments.

And what these discussions have shown is that the parties have a dispute about whether Microsoft's ongoing investments in OpenAI are relevant to discovery. And here, specifically Open -- sorry -- Microsoft's 2024 investments in OpenAI. And Microsoft seems to be taking the -- Microsoft is taking the position, at least so far, that 2024 investments and ongoing investments are irrelevant to discovery.

And we disagree. We allege claims against Microsoft for contributory infringement and for vicarious infringement as to OpenAI's ongoing infringement, and Microsoft's financial assistance to OpenAI, its ongoing investments in OpenAI are relevant to those claims.

So we think Microsoft should be ordered to produce documents in response to this RFP, and really to all the RFPs that have got these investment requests that focus on their ongoing investments. We don't think ongoing investments, in particular 2024 investments, should be carved out. So that is the dispute here. It's really a dispute about the timeline of the investments and which ones are and which ones are not relevant.

THE COURT: Let me hear briefly from Microsoft.

MR. BRIANT: Very briefly, your Honor. My apologies. Jared Briant on behalf of Microsoft.

Very briefly, your Honor. That is not the RFP that is

briefed in this motion. Specific request for production 123, briefed in the motion, is documents concerning whether Microsoft's investments in OpenAI and a number of years would affect OpenAI's nonprofit status, mission, or ability to engage in noncommercial activities.

Frankly, your Honor, OpenAI's nonprofit status is what it is. And from our perspective, document requests geared towards Microsoft for documents about whether our investments impact their nonprofit status or mission, I just don't think is relevant in this case.

That's the RFP that is at issue in this motion. This is another example of us having a discussion about an issue that is really not before the court. The issue Mr. Frawley is raising is something different that is not referenced here in the motion.

This is an RFP directed towards the impact of Microsoft's investments on OpenAI's tax status, which, again, is what it is. And, frankly, I don't think it's necessary or relevant, a document request geared toward Microsoft.

We can certainly continue to have the discussion on the broader issue with the news plaintiffs, but that is not what's before the court right now.

THE COURT: All right. With respect to RFP 123 then, what's the dispute?

There was an objection because it's not relevant?

MR. BRIANT: Absolutely, your Honor.

Yes, we've produced documents regarding investments in OpenAI. But this is asking for something different and more, in our perspective, it's documents about the impact, I suppose some sort of analysis they might have on what OpenAI's nonprofit status is.

THE COURT: This brings me back to the beginning of this conference when we were talking about hypothetically documents that one of the defendants might have talking about, you know, what might affect the other defendant.

I mean, wouldn't this stuff sort of just get produced anyway, or is it not -- is it going to get --

MR. BRIANT: The investments themselves are getting produced. Frankly, beyond that, what to look for, I struggle with that.

THE COURT: Well, then that's part of the meet-and-confer, right. If there is internal e-mails talking about the investments, you know, they are what they are. I can't imagine that it's in anybody's interest to screen things like that out.

So you're going to get them. Raise this later if it becomes a problem. I'm talking to Mr. Frawley.

As far as this dispute about the timeline and how investments and how it might relate to contributory negligence, I think I'm going to table all of that until we see what

Judge Stein says in the motion to dismiss.

MR. BRIANT: That's fair, your Honor. We can have this discussion then.

MR. FRAWLEY: Your Honor, can I make one really brief point?

THE COURT: Yes.

MR. FRAWLEY: OK. So, for the contributory -- so, the vicarious claim, Microsoft did not move to dismiss that claim. And for the contributory claim, they moved to dismiss a different version of the claim. They moved to dismiss the claim that says if a user infringes a copyright, then Microsoft is contributorily liable for the user's infringement.

They did not move to dismiss the claim that Microsoft is contributorily liable for OpenAI's infringement. So I appreciate your Honor wanting to see --

THE COURT: I get it. I'm also, in terms of the disputes that are bubbling up, I also do think that this is one that you may want to continue to meet-and-confer on.

MR. FRAWLEY: Thank you, your Honor.

MR. BRIANT: Thank you.

THE COURT: I'm not actually sure what that specific dispute about 2024 investments even is at this point. I'm not sure it's concretized to a point where I really need to look at it yet.

All right. I think on my list this brings us to

deposition coordination. Does anybody disagree?

Because, for me, I want to talk about deposition coordination last. It will be very short. I've been wrong before, but I don't expect us to be here more than another 10 or 15 minutes.

MR. DAWSON: Your Honor, Andrew Dawson for OpenAI. I did want to flag, we had a motion at docket 388, which is on page 20 of the chart in The Times case.

THE COURT: OK. Let me pull up that chart.

MR. DAWSON: And then I believe there is another one at 383.

THE COURT: OK. I may have not put them on my big list, because I think that they are either premature or I think you should meet-and-confer on them, but let me take a look.

Sorry. Where are they on the chart, page 20?

MS. HURST: Page 20 in The Times chart. Towards the bottom there is motion at docket 388, and then there is an additional one -- and I apologize -- on page 19 as well, there is docket 383. Page 20 is the docket 388. And then there is an additional one on page 21. The public filing is at 397. These are all kind of three in a row.

THE COURT: OK.

MS. HURST: Pages 19 to 21.

THE COURT: OK. I thought the documents, you were resolved, or it's no longer an issue. That would be 394, 397.

MS. SOLOMON: Your Honor, we are close to resolution on that. We're happy to continue to meet-and-confer.

THE COURT: OK. That's why I knew ...

OK. Let's go to 388. Oh, this is the search engine issue.

OK. I have them lumped under here and I didn't put numbers there.

This is Microsoft's motion.

MR. DAWSON: This is OpenAI's motion. It has a component of pricing information and then a component of search engine indexing and search engine copying.

THE COURT: OK. So, for price changes and bases for changes, why do you need more than what you're getting?

MR. DAWSON: So, your Honor, the reason we need more is that one of our challenges in this case has been to substantiate The Times' allegations of loss as we've looked at headline numbers, which The Times cites repeatedly in discussing its productions. Those headline numbers all reflect subscribers are going up, revenue is going up, profit is going up.

So obviously The Times has something else in mind to demonstrate the harm that it's alleged. So, RFP 200 is an attempt to fill in that gap, not to reinvest the wheel again to keep that theme going. If, in fact, there is the harm that The Times has alleged, we anticipate it must show up somewhere.

One place it would show up is when The Times is discussing market forces that might affect its prices.

If, in fact, The Times has felt under siege and has been suffering this harm, you would imagine that those perceptions would be demonstrated in these kinds of communications. They are not demonstrated in the headline numbers that we have received, so we need to find evidence of it somewhere else either to corroborate their allegation or to undermine it and to challenge it, which is, of course, our right.

I think it gets to a deeper issue that arises in a lot of RFPs, where The Times has been very, very resistant to producing communications. In many contexts, the response has always been, Oh, but you have the headline numbers. You have the high-level numbers. You have the noncustodial numbers.

And in many instances, that is enough, but in some of these critical issues involving harm, which goes both directly to damages and goes to market harm for fair use, there simply is no substitute for communications. And that is what we have not received.

And for RFP 200, communications about these pricing decisions are precisely where you might imagine these economic pressures being reflected. If the communications do not reflect pressure from OpenAI's offerings or other offerings, if it aligns with the headline numbers and seems to be business is

going great, all the numbers going up, that's obviously critical evidence for us.

And, in general, we just need to get deeper into the rationale and understanding of The Times' alleged harm.

THE COURT: All right. Let's hear from The Times.

MR. MUTTALIB: Thank you, your Honor. Adnan Muttalib for the *New York Times*.

I think that OpenAI's argument continues to show a misunderstanding of what our theory of harm actually is. Our theory of harm has consistently been that their products are causing users to less likely turn to The Times website in the first instance. It has nothing to do with our anticipated pricing, what our price might be in five years. It has simply to do with the fact that there is a reduced amount of traffic of individuals coming to our website which is leading to a decrease in advertising revenue, affiliate link revenue, and potential lost subscribers.

As a result, even if there were documents showing discussions about anticipated pricing, that would have no bearing on our actual theory of loss.

THE COURT: OK. Have you produced information about documents about traffic to the site.

MR. MUTTALIB: Absolutely, your Honor.

THE COURT: OK.

MR. MUTTALIB: Not only have we produced documents

regarding traffic, we have produced the reasons for the fluctuations in traffic regarding their products as well as other market forces. We have produced documents with relation to subscriptions, revenues. All those documents bear on our theory of harm. They have everything they need.

This is another attempt to try to get more documents on things irrelevant. As a result, we don't think they are entitled to them.

THE COURT: I'm not seeing pricing changes as something that needs to be produced right now.

MR. DAWSON: Could I respond to just one?

THE COURT: Very briefly.

MR. DAWSON: Nobody is misunderstanding the contention of harm. The fact of traffic to the site, what we are missing is the linkage between this alleged loss in traffic and some form of economic harm. Counsel just mentioned loss of subscribers, loss of advertising revenues. We do not see that in documents that we have received.

The high-level numbers do not show that loss, so there is a critical linkage. The Times alleges traffic is the problem, but it's not showing up as damages, it's not showing up as harm. We need to fill in that gap.

THE COURT: Wait a minute. What do you mean it's not showing up as damages or not showing up as harm?

MR. DAWSON: Counsel just mentioned that they have

produced, and the publicly available in The Times 10-K, things like recent records for advertising revenue, recent records for subscriber numbers, recent records for profits per subscriber. All those numbers are going up. There is no evidence in those numbers that this alleged loss in traffic is, in fact, leading to harm. That is the critical linkage that we don't have, and that is what we need and that is what RFP 200 is designed to help get us at.

THE COURT: No. I don't see how pricing changes has anything to do with that.

Isn't this an expert issue?

MR. DAWSON: Well, we need some evidence for the expert to analyze. At the moment, we just have a black box.

THE COURT: You have traffic numbers, right?

MR. DAWSON: We have revenue going up.

THE COURT: And you've got an expert who is going to talk about, you know, substantiate or not what Times' allegations of harm are.

Yes, Ms. Hurst.

MR. HURST: Your Honor, if I may, Annette Hurst from Microsoft.

Times has actually refused to produce a lot of the traffic data that we asked for, and that was the subject of the court's follow on ECF 344 orders that are now the subject of our Rule 72(a) objection. So it is not the case that The Times

has agreed to produce all of its traffic data.

MR. DAWSON: And that is certainly true. There is one RFP that The Times, OpenAI RFP cited repeatedly about analyses of traffic. But the granular requests, the requests Ms. Hurst is referring to, those have not been produced. What we have now is the black box and insufficient data to engage with that analysis.

THE COURT: Right. And you have a prior ruling from me that's up before Judge Stein as a Rule 72 objection, right?

MR. DAWSON: Well, we have a different basis for this one as well. So these are different RFPs. RFP 200 is not subject to that.

THE COURT: What is before Judge Stein?

MR. DAWSON: So there was a previous, at the last hearing, there was a market harm base motion. That is up to Judge Stein. What we had previously discussed with your Honor in October, when I appeared in front of you, we raised the issue of damages, and your Honor advised that it was not yet time for damages.

And we're now, obviously, deeper in discovery and we think now is the time. And this is precisely the kind of RFP that gets to the damages issues that were not, clearly were the focus of the prior briefing, but they are clearly a separate independent basis for discovery of these materials.

THE COURT: So, I'm still not seeing drawing the line

between pricing changes and proposed pricing changes and how that gets to the damages.

MR. DAWSON: So the idea is what we've been trying to find some category of documents where if it is true, as The Times alleges, that they have suffered all of these economic harms through diverted traffic, that there would be some conversation, some reflection of that harm.

And pricing discussions are one area where you might expect that to happen. If there is some competitive pressure, some issue, you might imagine discussions of cutting prices to increase subscriber rates. If there is no such pressure, if, in fact, OpenAI's conduct is irrelevant to The Times' financial well-being, you would imagine those conversations would not refer to OpenAI. They would not refer to any of those market pressures, or they might refer to entirely different pressures.

If pricing analyses are being driven by perception of post-pandemic realignment in reading habits, if it's the rise of social media, if there are entirely other factors, all of these things go directly to damages of whether the alleged harm, in fact, is traceable, first of all, whether it exists and, second of all, whether it's traceable to OpenAI.

THE COURT: OK. I see Mr. Muttalib is standing up. Why don't you respond.

I'll just let him respond first before I say anything.

MR. MUTTALIB: Thank you, your Honor. Adnan Muttalib

for The New York Times.

Respectfully, this misconception, this idea that we haven't turned over the traffic documents. I mean, as you know, your Honor, you have already --

Their attempt to relitigate this issue is improper and we have already dealt with it. We have turned over all traffic documents we are required to consistent with the court's order.

Second, with respect to the anticipated pricing, your Honor, I think we fully agree with your questioning, which is to say it's not the causation link between anticipated pricing and market harm. It is just our theory of harm is completely attenuated. I mean, it could be the case that anticipated pricing going up or down has nothing to do with the amount of traffic to our website.

We fail to see why that would demonstrate the type of loss theory that they need, especially given the fact that our complaint has no allegation that we have increased or decreased our prices which is has led to a loss of revenue. As a result, we don't understand how this category of evidence in any way is relevant to our theory of harm and them disapproving it.

THE COURT: OK. This seems to me an issue for, at most, expert discovery. It's still too soon. I'm not going to order that this, the pricing changes and the bases for changes and documents relating to them should be produced at this time.

All right. Search engine info. Again, my question

for OpenAI is, what more do you need and why?

MR. DAWSON: So, your Honor, on this one, again, I think it comes down, in part, to some of the custodial distinction, but I think at a higher level, obviously The Times felt the need in the complaint to set aside, to try to distinguish the kind of copying that search engines engage in, the kind of copying apparently The Times permits.

The fact that it was specifically carved out seems to demonstrate its relevance that there are obvious similarities here. Now, The Times argues there are differences to, and certainly there are. whether those differences are material or not, it is not enough for counsel to simply allege they are different. It's close enough, in fact, close enough in the kinds of copying, close enough that it was even raised, I think, last week in the hearing before Judge Stein.

Counsel for The Times went at some length talking about some of these distinctions and why those distinctions matter in this case and how they are different. But the fact that they are close enough to need distinguishing demonstrates that they are relevant. And it's not enough for counsel to try to simply carve it out through artful pleading.

We seek discovery based on those same allegations to assess some of the details that are not present in the complaint, things like the terms by which these things are permitted, if there are further details on the forms of search

engine indexing that are permitted or not permitted. These are the details that we lack and the details that might help illustrate what The Times contends is different about OpenAI's conduct and ways in which, perhaps, they are, in fact, similar. Not in every way, but perhaps in ways that matter.

THE COURT: I mean, a search engine index is not the same as the whole copyrighted content, right.

MR. DAWSON: I'm not sure that's entirely true.

Part of it is we don't know exactly the details, but search engines make copies of vast quantities of the data on websites in order to perform their functions.

THE COURT: Mr. Muttalib. I'm sorry I keep mispronouncing your name. Come on up.

MR. MUTTALIB: Adnan Muttalib, again, for The New York Times.

Your Honor, I think this one is patently obvious. The use of Times' works by search engines generally is obviously going to be different than the use of by defendants and their products. They literally conceded this last week at the motion to dismiss hearing when they opined on the record, stating that their products are meaningfully different from that of search engines generally. And as a result, we fail to see how this evidence is relevant.

But even taking them at their best and saying we have given some search engines access to our content no way bears on

whether we gave Open AI access to that content. Now, it would be like me inviting you over for a dinner party and then these guys showing up and saying, we are also lawyers, so we are allowed in, too. It's obviously different. It is meaningfully not the same thing. This evidence is, again, just another attempt to try and get information that is not relevant to their claim.

THE COURT: Mr. Dawson, I do want to understand why, what claim, what claims or defenses this has to do with.

MR. DAWSON: So I think, in part, it would go to -- I mean, the precise terms by which The Times permits search engines and the factors considered by The Times, it appears to us that The Times believes that such uses are fair or fair uses. It is not our understanding, but perhaps we're wrong, that there are, in fact, commercial agreements between every search engine in the world and the Times to permit this kind of behavior.

So, in our view, it is close enough to what OpenAI is accused of doing, and yet The Times believes this conduct is fair and clearly believes OpenAI's is not. Nobody is contending these are the same. Obviously, there are some differences.

THE COURT: But, I mean, isn't the fair use analysis, isn't to say that we're just like the search engine or we're like the search engine is enough?

I mean, you have to prove your fair use defense on your own, on OpenAI's own use, not on what a search engine is doing or not doing with The New York Times' content, right?

MR. DAWSON: We certainly intend to do precisely that, but one aspect of the question will be, first of all, has The Times taken different positions over time, is there some inconsistency. If there is a witness on the stand testifying about one thing that search engines do, if it turns out that The Times' posture with regard to search engine indexing was different previously or changed over time, these are all things that we think are within bounds. Obviously, it's not the whole story.

THE COURT: I don't agree. I don't agree. You're not getting it. You're not getting it.

MR. DAWSON: Thank you, your Honor.

THE COURT: All right. I'm not seeing the relevance there. If you want a clear ruling that you can appeal, I'm not seeing that what The Times does with search engines and how it's allowed certain aspects of its content to be used for search engines to redirect back to The New York Times, by the way, is relevant to what The Times has allowed or not allowed OpenAI to do with its content.

OK. All right. Deposition coordination.

MS. SOLOMON: Your Honor, we have a remaining motion on OpenAI's RFPs 36, 37, and 39. I can provide argument on

that if the court wants.

THE COURT: Point me to the ECF numbers.

MS. SOLOMON: It is at docket 383 is the original motion, and 401 for the response. And the chart, it's the very bottom of page 19.

THE COURT: OK. I'm getting there.

MS. SOLOMON: I neglected to state my appearance. Sarah Solomon for OpenAI.

THE COURT: I don't think RFPs, requests for admission, are the proper vehicle here. Has The Times actually served responses to the requests for admissions?

MS. SOLOMON: No, your Honor. The Times has said that it's unable to respond to the RFPs.

THE COURT: I think my question is, has there been a formal response to the requests for admissions?

MS. SOLOMON: Yes, your Honor. There's been a formal response neither admitting nor denying.

THE COURT: OK. Then that's your response.

MS. SOLOMON: May I respond just briefly to that?

THE COURT: No. No. We're done. I already said at the beginning, this is not a proper vehicle for an RFP. And, you know, you have a lot of discovery ahead of you. You could explore these issues in deposition with other documents and the like.

MS. SOLOMON: Thank you, your Honor.

THE COURT: Now are we on to deposition coordination? Yes? Maybe.

MS. GARKO: Yes, your Honor.

THE COURT: OK.

THE DEPUTY CLERK: State your name.

MS. GARKO: Sheryl Garko on behalf of Microsoft.

THE COURT: Should we take the easier or the most pressing issues first, which might be the 30(b)(6) custodial 30(b)(6) deposition?

We ruled on those. We ruled on those. Nevermind.

OK. All right. Deposition coordination.

Let me start by saying that the issues, as I see it -- and this is now going to apply to the Authors Gild cases as well as the news cases -- but I see the issues in the news cases as being wider and broader. So I'm not going to force you to do a one-size-fits-all for the Authors' cases here. The Authors' cases or case in the Northern District of California and the news case here.

What my overall, overarching guideline here is, coordinate as best you can with the Authors' cases here and the Authors' cases here and the Authors' cases in the Northern District of California. OK. It seems to me that there can be a lot of efficiencies gained by coordinating for those. But I'm not going to tell the news cases that they have to go to the same level of coordination.

OK. I do, however, hope that the news cases and the news plaintiffs will coordinate among themselves.

OK. Starting from there.

MS. GARKO: Well, if I may address that, your Honor, because I think one of the threshold issues here is why that coordination in New York is actually necessary and why it's necessary to do it now.

So, the threshold issue here, your Honor, is we are at the deposition stage. It's now open. Custodial depositions have started. Fact are soon to happen, likely before this next stage. We really do need some guidance here from you on the rules and the limits so everyone knows what they are doing, what they are playing within.

And why coordination is necessary within New York, your Honor, is because of the overlap of issues here. I agree with with you entirely that there are additional issues in The New York Times case, but that doesn't mean there aren't overlapping issues and significant overlapping issues with the class case in New York.

If you think about this, your Honor, like a Venn diagram. You have the issues with respect to training, which between the class case and the news case are 100 percent overlapping. With respect to Microsoft there, for example, your Honor, all of the custodians in the class case are custodians in the news case. All of our folks on the initial

disclosures in the class case are on the initial disclosures in the news case. There is a one-to-one correlation. All of the documents, your Honor, that have been produced in the class case have been produced in the news case.

THE COURT: OK.

MS. GARKO: There is significant overlap. Then yes, you have this additional issue with respect to search outputs that is added on for The Times. But with respect to that overlap, your Honor, it's going to be the same witnesses, it's going to be the same documents.

And having those witnesses have to testify more than once, having those depositions uncoordinated, is just going to create an enormous amount of inefficiency and is not consistent with the directives of Rule 26 to avoid duplication. So that is why we believe that it's really important to have help here.

THE COURT: OK. I'm not saying there should be coordination. Talk to me about what the dispute is.

The way I read this was, you're having disputes about capping hours for depositions and defining Apex witnesses, but I don't even know what the Apex witnesses are. So I'm not sure we're there yet, and I'm not sure that a one-size-fits-all for every witness is appropriate.

MS. GARKO: I think, your Honor, with respect to the threshold issue is right now the proposal for Microsoft on the table is more formal coordination within New York. There is no

proposals that come from class plaintiffs that offers any coordination with the news plaintiffs, and there is no coordination proposal that is preferred from the news plaintiffs that involves the class plaintiffs.

So, right now they are seeking to operate entirely independently and not coordinate with each other at all, so that is where we believe that is not going to lead to those efficiencies.

THE COURT: In other words, for example, Microsoft wants, say, it's the same witness, you don't want them deposed twice on two different schedules.

MS. GARKO: Correct, and have duplicative questions asked of them, where the issues are going to be identical for many of those witnesses.

THE COURT: Why can't you make them available, say, for two separate days or two consecutive days, right, and do all of the overlapping issues as much as you can the first day.

MS. GARKO: Your Honor, we have proposed that type of proposal, trying to have the plaintiffs coordinate and do that. They have not offered that to us at all.

THE COURT: OK. Plaintiffs, talk to me. This still seems to me like something you ought to be meeting-and-conferring about more, but I'll hear from you on that.

MS. GARKO: And just respectfully to that point, your Honor, we have been really trying on this. We have been

talking about it since August. I personally have spent dozens and dozens of hours trying to get there. I think, unfortunately, there are just fundamental disagreements between the parties about how this should be approached.

MR. SAVAGE: Yes, your Honor. I'll go to the podium.

THE COURT: Let's hear from Mr. Savage.

MR. SAVAGE: On this issue of the coordination among the New York cases, I'm not sure there really is a dispute here, at least from the news plaintiffs' perspective. We have put forward two proposals. Our preferred proposal is a news-only proposal, which is consistent with what your Honor proposed or suggested that you were inclined to order at the beginning.

Even under that proposal, we would certainly welcome the opportunity to confer with defendants and with the class plaintiffs to schedule depositions consecutively, which is, I think, what your Honor just suggested. And so the news-only proposal is what was attached to our response as Exhibit 1.

We also said that we are open to an SDNY all-in proposal as well, and that would be a more formal sort of coordination. We don't think that's the best approach, including because there have been -- we have gone back-and-forth over this several times, but this issue of cross-production of documents.

The defendants have said that they want coordination

among all of the SDNY cases, but are not willing to agree to cross-production of documents across all of the SDNY plaintiffs. Our view is that they have to go together. But that's fine if they don't want cross-production of documents across all the SDNY cases. Then fine, we will just do news only. But, either way, we are certainly willing to do what your Honor suggested, which is schedule depositions, like, in a coordinated way.

THE COURT: Let's get really, really practical in the next five minutes.

All right. Go ahead, Mr. Nath.

MR. NATH: Your Honor, I think what you suggested -- and I don't think there is a lot of dispute here actually -- what you suggested --

THE COURT: That's why I saved this issue for last. I didn't think it would take that long.

MR. NATH: -- is right. The class plaintiffs, the Authors' plaintiffs in SDNY, should coordinate informally with the Authors' plaintiffs in the Northern District of California, because there are a lot of overlapping issues. The one issue that is not overlapping between those two cases is Microsoft, and we are totally fine with scheduling depositions back to back.

I think what that leads to is the dispute that currently exists is whether -- is about hours caps, like your

Honor said. Hours caps about whether, if there are two depositions in the news, you know, news case deposition scheduled before a class deposition, or vice versa, whether there is some sort of shorter time that the court should put on, a time limit for that deposition, that the court should impose prematurely we think.

So we think the informal coordination among SDNY, particularly for Microsoft depositions, makes sense. Informal coordination between the two class cases in NoCal and SDNY makes perfect sense, and we're perfectly prepared to proceed that way.

There are other issues related to hours cap that I think would be incredibly helpful to resolve here. I agree with Ms. Garko that we all put a lot of time in this, and we could use some clarity on caps.

THE COURT: This seems to me like the type of unproductive settlement conference where parties are just throwing numbers back and forth at each other and maybe you just need somebody to tell you.

But what I took from what Mr. Savage was saying was there is an open to an SDNY all-in proposal, but there are concerns about cross-production. I feel like this is really, I think, up to Microsoft.

How do you want to handle that?

Go ahead.

MS. GARKO: Your Honor, there really is no issue with respect to cross-production.

The fact is that, with respect to Microsoft, 100 percent of the documents that were produced in the class cases have been produced to the news plaintiffs. So it's odd that they are even raising this as an issue, because they have received every single document that the class plaintiffs have received.

Because there are additional issues that are only relevant to the news cases, they received additional documents --

THE COURT: I see.

MS. GARKO: -- where there were custodians relevant to the class plaintiffs that were relevant to news. The class plaintiffs got those documents. The reason why class plaintiffs have gotten fewer documents is they have one less issue.

THE COURT: Right.

MS. GARKO: So there really is no issue here with respect to cross-production at all in terms of coordination.

THE COURT: OK. So, for the class plaintiffs, have you all been talking about hours caps just for the class plaintiffs' deposition, or have you been talking about hours caps trying for all in?

MS. FARKO: So we have been talking about all in,

your Honor.

Importantly here, the threshold is, right now under the federal rules, we are numbers caps. We would like that converted to an hours cap, and then figuring out what that hours cap is. And we are at --

THE COURT: You've got numbers caps and presumptive hours caps. OK. I'm not going to -- so unless I hear some other better idea, I want you to think about whether it makes sense to have each Microsoft witness appear -- and taking your point about the Venn diagrams, right -- starting with the overlapping issues deposition, maybe that goes presumptively for seven hours. And then for the additional depositions, which for the additional issues that are unique to the news cases and that involve the news documents that you have produced to the news plaintiffs, that there is some additional time of deposition for each witness.

MS. GARKO: I think, your Honor, the witnesses that are going to be relevant to just the news cases, I don't actually foresee being relevant to the class plaintiffs. I think the issue is for just news. So the issue search outputs, that is just for news, isn't going to be relevant to the class plaintiffs. The issues that are relevant to the class plaintiffs on training will be relevant to the news plaintiffs.

THE COURT: Yes.

MS. GARKO: I think that is the number that was in

dispute in trying to solve.

THE COURT: You were talking about the common issues.

MS. GARKO: Correct.

THE COURT: Time. I said, how about seven hours.

MS. GARKO: Your Honor, we would have no issue with seven hours on those issues.

THE COURT: For that plus, or are you saying that that is not even going to be enough for...

MR. NATH: Your Honor, the issue may be common to some extent, but there are differences in the documents that are at issue.

THE COURT: You're right.

MR. NATH: For example, there are books datasets, right. When we are talking to either a Microsoft witness or an OpenAI witness about training, the class plaintiffs will be focusing on the books datasets and other datasets that contain books, while I expect that the news plaintiffs will be focusing on different datasets.

So we do think that that seven hours all-in for a deposition that may involve training and only training, even though that issue overlaps, isn't necessarily going to be sufficient.

THE COURT: Why can't you do what litigants in virtually every other case have done, is get to your list of witnesses and talk about how long each witness you think is

going to take.

Why are we here?

MS. GARKO: Your Honor, I think the threshold issue is the ultimate hours cap.

THE COURT: I'm not doing -- no, we're not doing an ultimate hours cap. We are not doing an ultimate hours cap when you don't have a sense of who the witnesses are and how long each witness is going to take, because that is just silly and it's not tailored to the needs of the case and the needs of the availability of each witness and it doesn't --

No. No. That's a hard no.

OK. Have you talked about a list of witnesses and whether some witnesses might be a two-day deposition, some witnesses might be a four-hour deposition.

Why can't you do that?

MR. NATH: Your Honor, I think we could, from the class plaintiffs perspective. I think we could do that. I think every party here is of the view that an hours cap would be helpful to guide us as we go into discovery and go into depositions, and we are far apart on the hours cap, but not that far apart.

THE COURT: You're talking about a total hours cap?

MR. NATH: Talking about a total hours cap from the plaintiffs' side per defendant, and the defendant have proposed an aggregate hours cap to share between them for each of the

separate coordinated cases.

If we are turning to another issue, I do want to address Ms. Garko's argument about cross-production. We think that is a sort of separate issue that I want to address briefly. The cross-production issue with respect to Microsoft I would like to set to one side.

From the class plaintiffs' perspective, we -- as the court had suggested we do and as Judge Elman in the Northern District instructed the Northern District of California plaintiffs to do -- intend to coordinate with that case to try and schedule depositions back to back and minimize the amount of time on the record and coordinate with those plaintiffs for depositions.

Right now we can't do that because we have a protective order that governs this action, they have a protective order that governs that action, and we can't share documents that OpenAI has produced to us, which is why we're asking for this court to order OpenAI to cross-produce the documents that they have produced in the Northern District of California action to us.

Because as a practical matter, if I show up to a deposition that the Northern District of California plaintiffs are taking, my understanding is they will have different Bates numbers. I would like to know whether the document that they are questioning the witness on is actually a documents that has

been produced to us, something that we can print out and use. And even just keeping track of the different documents in the case, it would be incredibly helpful. If OpenAI's goal is to minimize the burden on the witnesses and minimize the time of the record, it would be incredibly helpful and efficient to actually have cross-production of documents from OpenAI between the Northern District of California and the SDNY case.

THE COURT: You just brought up a new issue. All right. And the problem is that the cross-production issue needs to be resolved because it might impact on how long you need to take with depositions.

MR. NATH: It impacts the ability for us to actually and formally coordinate with the Northern District plaintiffs. Right now we can't even share documents that OpenAI produced in this case with the plaintiffs in that case.

So if we're going to prepare for a deposition, say we're both deposing the same witness, I would like to be able to communicate with those plaintiffs in advance of that deposition. OpenAI's suggestion is that we should provide them the documents in advance and ask what documents we should be able to share with the Northern District of California plaintiffs.

We already discussed the element of surprise. We prefer not to give a roadmap for the depositions before every deposition in the case.

THE COURT: All right. Mr. Slaughter, you're standing up.

MR. SLAUGHTER: Thank you, your Honor. James Slaughter on behalf of OpenAI.

Just with respect to this cross-production issue, your Honor. You know, there is a suggestion here that there hasn't been cross-production, and that's just false. More than 90 percent of the OpenAI documents have been produced in all of the cases. That is the class. And the news cases, and has Ms. Garko noted, there are some differences because the news claims are broader, there is more custodians in those cases. So there will be differences.

But this is, as you noted, not an issue that has been teed up for now. We can meet-and-confer on it. There is, obviously, very easy ways to address it, including, most simply, if there is a witness who is being deposed, making sure that they have -- that there's been the same documents have been produced both in the class cases so that everybody is working off the same dataset, so to speak.

The suggestion, A, there hasn't been class production is not fair or appropriate, and that we can't handle this on a case-by-case basis as we schedule these depositions. That's the way to do it, is to make sure that, you know, if you have a deponent coming up, everybody has got the same set of documents.

THE COURT: Right. That's really on OpenAI, right. You're the ones that have been producing to everybody.

MR. SLAUGHTER: Yes.

THE COURT: OK.

MR. NATH: So, your Honor, the problem that that doesn't solve is it doesn't solve our ability to actually share documents with the other plaintiffs that we are supposed to coordinate with. Presumably, we are also going to be coordinating about witnesses we are going to depose. Instead of doing this piecemeal and having more meet-and-confer efforts over cross-production, if 90 percent of the documents that have been produced in our case produced in the Northern District of California case, or vice versa, OpenAI should just produce that other 10 percent.

There are enough overlapping issues that we don't understand why this should be a difficulty. And it also constrains our ability, for example, to actually share a document. If we think that both cases should depose a particular witness and want to explain why, we can't do that given the protective order in our case, and have those open communications with the Northern District of California plaintiffs.

THE COURT: Let's get back to hours caps and number.

Have you talked about the number of witnesses?

MS. GARKO: Your Honor, we have not exchanged an

anticipated deponent list or anything like that, and that's part of the reason we have been struggling...

Your Honor, that exchange has not happened yet with respect to the list of specific witnesses. And that's one place where we've been struggling with plaintiffs' proposal because it is largely just based on custodian numbers, which we know there is also more custodians than folks that get deposed.

So we don't have a great sense of where their cap numbers are coming from. We set out in our briefing how we calculated from the defendants' side to come to the caps we came to, which is tied to the federal rules. But candidly, your Honor, I don't know where their proposals come from, because we don't have a sense of how many witnesses they actually anticipate that they are going to take depositions from.

On our side, your Honor, defendants know we want to take abbreviated depositions of all of the named class plaintiffs. There is 32 of them. We want a seven-hour deposition of the Authors Gild, so we have asked for that be hours. With respect to the entities, what we tried to do there, your Honor, is treat everyone equally. So instead of the standard each side is subject to ten depositions, we split that by party instead, and then just applied it across so everyone gets that equivalent amount of hours that they are

subject to deposition of.

THE COURT: OK.

MS. GARKO: Which recognizes, your Honor, the increased complexities, but also doing it this way, we can actually get through all of the depositions that we need to get through.

THE COURT: OK. So you can choose, for example, to spend more time with one witness and less time with another.

MS. GARKO: That's the benefit of the cap, your Honor, just flexibility there. If it's a four-hour witness, great. If it's more...

THE COURT: Mr. Slaughter.

MR. SLAUGHTER: Thank you, your Honor.

To the extent it would be helpful, your Honor, I would just direct your Honor's attention to ECF 430-1 in The New York Times case, which is a chart that the defendants prepared that has the one page, the varying proposals, so your Honor has it.

I join with how Ms. Garko just described the issue, and I would further say, your Honor, that we're trying to be as consistent as we can with Rules 30 and 26.

THE COURT: OK.

MR. SLAUGHTER: There has to be a reason to exceed the ten depo limit. We have explained why that -- what that reason is, but --

THE COURT: Oh, you're getting more than ten

depositions. That's not the issue.

MR. SLAUGHTER: Right, so it's not the issue. So the question becomes, under Rule 26 you have to prevent duplicative and cumulative discovery. And the plaintiffs' proposals all assume three separate plaintiffs uncoordinated and effectively unlimited with four or 500 hours of deposition time, that is assuming that, you know, it's a recipe for disaster.

What we are suggesting is, enter in the order that we have asked. If it ends up not being enough, we can come back. And we have taken some depositions and see and they can make a showing about why they might need more hours.

But starting at their level would just, sort of, not be consistent with the requirements of Rules 30 and 26.

THE COURT: Mr. Nath, you have one minute.

MR. NATH: I don't think you're going to be surprised. We disagree with what our opposing counsel has said here. So on hours limits, first of all, I think the court cannot just enter the order the defendants proposed because --

THE COURT: I'm not going to. I'm not entering any order, so let me just hear from you. Just a quick response.

MR. NATH: Yes. Let me talk about hours cap.

Our proposed hours cap for the plaintiffs taking depositions of the defendants and third parties is 225 hours, and I can explain exactly how we got there.

THE COURT: No, we don't need to.

Here is what I'm going to do. We're coming up on one o'clock. I thought we would be done before now, but I'm like Charlie Brown with the football. I think we can be more efficient.

Here is what we are going to do. January 31, 2:00 to 5:00 p.m., we are going to have a settlement conference to talk about deposition proposals.

OK. So every party, or group of parties, send one attorney or maybe two. OK. You can have other attorneys available to consult, but I really want the attorneys who come on the 31st to have the ability and to have the authority to make concessions, to talk to me.

I want the people who can actually talk to me about particular depositions, talk to me about why these issues are particular issues. This is an unusual case. It's a unique case. I get that. I'm going to take another look at 430-1, which I think has --

Mr. Slaughter, you said that has the competing proposals?

So I'm not going to ask --

MS. GARKO: We have copies of it, your Honor.

THE COURT: No, that's OK. I'll look at it electronically, too. I don't want you to submit anything in advance.

I want the people who can talk about your deposition

planning, to talk about what it is you might need, to be able to talk actually realistically about potential witnesses, and we're going to try to figure it out.

I'm not going to wear my robe. I'm actually going to have you all start talking first. I don't -- I'm not sure I'll be there for the whole three hours, but we're going to have a session. It's going to be confidential, and we're going to try to see if we can hammer some of this out.

MR. SLAUGHTER: Thank you, your Honor.

MS. GARKO: Thank you, your Honor.

MR. NATH: Appreciate that, your Honor.

THE COURT: Anything else?

MR. CROSBY: One small thing.

THE COURT: Yes.

MR. CROSBY: And it's not a discovery dispute, I swear.

THE COURT: OK.

MR. CROSBY: Just for clarification about the sealing procedures that we have been going through.

THE COURT: Yikes.

MR. CROSBY: There was an order granting the most recent spade of sealing motions. I think there is maybe a disconnect between the process that we have provided in the proposal order entered in the case and your Honor's individual practices. Because the way that the protective order works is,

if you're submitting another parties' designated document, says you just submit it. And the only reasoning you give for why you're submitting it under seal is because the other party designated it. Then the designating party responds, and then they have to justify why it should remain sealed or say we don't think it should be sealed.

And that's what we did, and then your order -- this was in the docket 378 with respect to sealed ECF 370 -- we said we don't think that this thing that they sealed needs to be sealed, and your Honor granted the sealing motion entirely saying that we hadn't justified unsealing it, which is kind of not the way the burden is supposed to be laid out.

I think your Honor said that you wanted more description of what was going to be sealed in the moving papers. But, again, that's not how the protective order we entered into works, because it's really the responsive party who is the one with the interest in the sealing.

THE COURT: Yes, I get that. I think additionally, though, there were -- I'm going on memory -- but I think I did do an analysis, on my own anyway, on why something ought to be sealed or not, does not need to be sealed.

This almost sounds to me, though, like a motion for reconsideration, actually.

MR. CROSBY: We obviously said in our opposition, we didn't care about that part being sealed. We don't.

THE COURT: Then I'm OK with it.

MR. CROSBY: OK. I just want to make sure we're all clear about what the process is.

Is the process that we've been following the correct one?

THE COURT: Yes.

MR. CROSBY: Thank you, your Honor.

THE COURT: OK. I think what you're referring to, there might have been -- we might have had some trouble. Because there was one motion, one sealing motion maybe that didn't refer to all of the documents, so that is why we were a little confused and we had to go back.

MR. CROSBY: Thank you for the clarification, your Honor.

THE COURT: Keep the ECF numbers in.

All right. I request the parties order a copy of the transcript. Share the cost, 50/50 on each side of it.

I will see some folks on the 31st for our deposition coordination settlement conference. I think, right now, all we need is a status letter on the 13th. We'll take it from there.

All right. Thank you very much.

We are adjourned.

ALL PRESENT: Thank you, your Honor.

(Adjourned)